DOKTORI (PH.D) DISSZERTÁCIÓ

A hamisvélekedés-tulajdonítás és a nyelv fejlődési kapcsolatai

False belief understanding and language: developmental relations

HAHN NOÉMI

Témavezető:
DR. GYŐRI MIKLÓS, docens

Eötvös Loránd Tudományegyetem, Budapest
Pedagógiai és Pszichológiai Kar

Pszichológia Doktori Iskola
Doktori Iskola vezetője: Prof. Hunyady György
Kognitív Pszichológiai Program
Doktori Program vezetője: Prof. Czigler István

2009.

Table of Contents

Acknowledgements

I would like to thank many people for their contribution to this Dissertation:

First, I would like to thank the children who participated in the studies and made it possible to explore these questions through their development. I am also grateful to their parents and the kindergartens, who made the testing possible.

I am especially grateful to my supervisor, Dr. Miklós Győri, who directed my attention towards developmental cognitive psychology and more specifically, to the development of theory of mind and language. I would like to thank him for all his exciting classes, useful comments and criticism during the past years. Working with him clearly shaped my thinking about these topics.

I am grateful to Anna Balázs, head of the Autism Foundation, Budapest and to Katalin Rosta, the head of the ELTE Special Kindergarten and Early Development Methodology Center for letting us perform the tests at their institutes.

I would like to thank Barbara Batta, Anikó Dósa, Pintér Ágnes, Eszter Sajó, Krisztina Stefanik and Zsuzsa Várnai for their help in testing the children and also for their useful comments.

I am very grateful to András Fiser for his endless patience and tolerance and for his support and comments.

I thank my parents Oszkár Hahn and Oszkárné Hahn and my siblings Katalin Hahn and ifj. Oszkár Hahn for their constant support.

Last but not least, I thank Piroska Korsós, Bernadett Samodai and Gabriella Szenderák, for their help in the official process of the Dissertation.

Chapter 1. Introduction

The developmental relation between language and false belief understanding or, in a broader aspect, language and theory of mind has been a flourishing field of research of developmental psychology. Approximately 324 studies were published (Milligan, Astington, & Dack, 2007), most of them within the last 10 years, which means one study per eleven days. This remarkable interest in the relation between these two (most probably) human specific capacities is partially due to its implications to several research areas, including atypical development, evolutionary psychology, (developmental) cognitive neuropsychology, cognitive science, anthropology and philosophy. The current Dissertation focuses on the developmental aspect of false belief understanding and language acquisition in typically developing children in the kindergarten age and also focuses on some special populations of children.

The question regarding the developmental relation between language and false belief understanding is not whether this relation exists or not, but rather the nature and direction of it. The title of a recent book "Why language matters to theory of mind" (Astington, 2005) also mirrors this attitude. However the lack of such relation was a real possibility, for instance, the same authors did not support this idea 10 years ago (Astington & Jenkins, 1999).

Following the state of the art on this issue, the Dissertation aims to gain more insight on disclosing the nature of this developmental relation by focusing on different aspects of language. Although the broader question refers to the relation between ToM and language, the Dissertation focuses only on the specific aspect of ToM; false belief understanding, the testing of which became the 'litmus test' of ToM after Dennett's famous paper (1978). However, the test itself clearly does not reflect the complexity of ToM capacity, as it will be discussed in Chapter 2. As we will also see, one of the major criticisms of studies testing this relation is that false belief understanding was accessed with a verbal test, which leaves open the possibility that whatever relation is found is due to the verbality of false belief tests (FBT), and does not reflect the real relation (e.g. Astington & Jenkins, 1999). That is why the first and second studies of the Dissertation (Chapter 5) focus on the development of a new nonverbal FBT. One of the language aspects, the role of which in FB understanding created active arguments, is sentential complement. De Villiers radical/strong statement of an

essential, causal relation between complement syntax and FB understanding motivated a number of studies to test this phenomenon. In *Study 3* I will also follow this line of research, however unlike the studies before, I will access FB understanding with a nonverbal FBT. Moreover, not only the original version of the de Villiers complement hypothesis will be tested but in *Study 4* also its modified version combined with word-learning. The idea comes from Happé and Loth's (2002) study, where they found that children passed FBT earlier if it was combined with a word-learning task, thus I applied this paradigm to the complement task too to test further hypothesis regarding de Villiers statement. Some of the hypotheses of the Dissertation were also tested on atypically developing children; on children with developmental language impairments (DLI) and on children with autism spectrum disorders (ASD). The relevance of these special populations is to test the relation on children when one of these abilities is impaired by definition and to explore whether and/or how it modifies the other 'intact' or not necessarily impaired ability. The implication of the studies both on typical and atypical children regarding the cognitive developmental models will be discussed.

Chapter 2. Theory of mind

Maybe due to the relatively young age of the concept of naïve theory of mind (first published in 1978) there are still many uncertainties about it. This is partially due to the fact that it can refer to more than one phenomenon. In the first usage of the term by Premack and Woodruff (1978) theory of mind referred to the *ability* to attribute mental states and to predict and interpret our own and others' behavior based on these mental states. The term however, in a more restricted way, can refer to a specific cognitive mechanism, to a so-called mentalizing *mechanism,* which is a "prerequisite for the attribution of mental states and a (necessary, but not sufficient) prerequisite for an implicitly held theory of mind" (Frith & Frith, 2000). The third usage of theory of mind refers to the *research area,* which investigates the development of these abilities (Astington, 2005). Abilities, theoretical explanations and research area are clearly different phenomenon, but it is also obvious that hey are liked together. Perhaps, that is the reason why the term has been used interchangeably. A few researchers (e.g. Hobson, 1991; Nelson, 2007) even refuse to use or attribute "theory of mind" to children, arguing that it is misleading even as a metaphor (for instance Nelson us the term "community of minds"). In the Dissertation, however I will stand for the approach (allaspontot kepvisel) that believes "theory of mind" is a useful term, but one has to precise in which aspect(s) of it we refer to. The term of theory of mind was also criticized because it implicates that what is in our head is a theory. But as we will see in Chapter 2 there are alternative assumptions and therefore it is still not a decided issue. Probably that is why there are so many synonyms to this term. Frye and Moore (1991) chose the term folk psychology or commonsense psychology instead of theory of mind they argue that "the child is not really developing a theory in anything like its scientific sense but rather a way of thinking and talking about self and other that involves mental states" (Frye & Moore, 1991, p.1). One of the most commonly used synonyms is mindreading (see e.g. Kiss, 2005), which comes from comparative psychology literature used e.g. by Whiten (1991) for primate social intelligence, but since mindreading in the everyday is also associated with telepathy the term did not gain widespread usage. Another frequent term in the literature is mentalizing ability, Happé and Frith use this term referring to the ability to impute mental states. In addition there are many other terms, which are closely related to the term theory of mind and often used (sometimes not appropriately) interchangeably with it e.g. perspective-taking, metacognition, social cognition, belief-desire reasoning, intersubjectivity, intentional stance etc. In spite of these uncertainties and ambiguity even after 30 years of

history of the concept of theory of mind, it is still a useful term and the most widespread one. To handle the disadvantages and ambiguity of the term, I will clarify the specific aspect(s) of the theory of mind I am referring to in each part of this Dissertation.

2.1.The nature of the developmental relation of language and ToM

As I mentioned earlier, the different theoretical models suggest different relation between the two developing abilities, however as we will see many times these models are not unitary. There are two major approaches regarding this relation. The first suggests that there is no special role of language in the development of theory of mind. However, one can get to this conclusion from various perspectives about the nature of ToM. Probably the most obvious supporters are the nativist modularist theorists. Since the approach suggests that ToM is innately specified and develop in a separate module, language is not necessarily needed for its proper development. It is important to note that by assuming the modular architecture of the mind, one can suggest a very different relation between ToM and language. Sperber (2000) and Sperber and Wilson (2002) for instance propose that there may be more than one theory of mind mechanism (ToMM)—and that the communication system may possess its own dedicated metarepresentational competence: '. . .the recognition of communicative intentions might be a biologically differentiated and stabilized sub-system of human naive psychology' (Sperber, 2000 p. 133).

The second assumption clearly denies both the innatness and the modularity of ToM, moreover even its domain specificity. They suggest that ToM abilities rest on domain general cognitive processes, and language is needed only to implement these cognitive processes. What these domain general processes are, that is different in the different approaches. For instance Riggs & Simons (2005) stresses the relation between general reasoning and theory of mind, where language is only needed to understand complex and conditional sentences, while Frye, Zelazo, & Palfai (1995) emphasizes other domain general processes such as cognitive complexity and control. Other researchers, such as Gopnik, Wellman (1994) and Perner (2000) still denies nativist modularity, they argue that the role of language is just a natural way of providing children with the information they need to build up, or construct a theory of mind.

Finally, regardless of the theoretical models, it is possible that even if we find such a relation between these two abilities it is only a by-product of the verbality of theory of mind tasks (Chandler, Fritz, & Hala, 1989).

On the other side, however, there are the theorists, who believe that the relation between the two abilities is fundamental and causal. But this approach is also far from being unitary. The major question is whether there is a specific aspect of language that play a special role in ToM development or language in general has this special role. And if there is such an aspect what it is? In Chapter 3 we will see a couple of possible language aspects that could play this special role.

2.2. The development of theory of mind

Although the Dissertation focuses on a particular level of theory of mind development, namely the false belief understanding, the earlier manifestations of this ability will also be discussed, because in case of a developing ability the whole process of development should be in focus, especially if we want to get closer to its relation to another developing ability (in this case to language). But it also has to be admitted that since early social cognition has been one of the most flourishing research field, its literature is tremendous, so I am going to focus only at the most relevant aspects of it, concerning the topic of the Dissertation.

2.2.1. Possible precursors of ToM

Although the Dissertation focuses on developmental relation of theory of mind and more specifically on false belief understanding and language, children/infants are clearly engaged in rather complex social interactions at much earlier than 3 or 4-yeas of age, when they start to pass FBT. The literature of social understanding in infancy is very diversified, so I am only going to discuss the further possible precursors of theory of mind: joint attention, pretend play and imitation.

2.2.1.1. Joint attention, Imitation and Pretend play

Joint attention is the capacity to coordinate attention with others in order to share or direct others' attention to a common point of reference (Mundy & Newell, 2007; Carpendale & Lewis, 2006). Joint attention is a complex term, involves both the initiation of these behaviors by the infant and the responding to these behaviors initiated by another person[1]. The

[1] It has been a question what the relations are among these different behavioral manifestations (e.g. Mundy &Newell, 2007; Tomasello et al., 2005)

behavioral repertoire of joint attention can vary e.g. gaze following, pointing, just like their function e.g. assumed attention sharing in proto-declarative pointing or sharing emotional state about a third object in social referring. However there are different views when these behaviors appear first, most of them does around 9-12 months of age (Carpenter, Nagell, & Tomasello, 1998). Joint attention is probably the most well known precursor of theory of mind, the theory was first suggested by Baron-Cohen (1989, 1995) then later Tomasello (1995) has presented a theory according to which the emergence of joint attention skills is an evidence for the infant's emerging understanding of others as intentional agents and that this social cognitive ability is the basis of a later developing theory of mind. Actually Tomasello changes the term to post-cursor and therefore the focus of the train of thought, suggesting the focus should be joint attention, since it is a fundamental skill to engage in the social world, and everything that comes later – including theory of mind – depends on it and therefore is less important (Tomasello, 1999)

The next ToM precursor candidate is imitation. Gopnik & Meltzoff (1993; Gopnik, 1996), repeatedly argued that in humans, the innate ability manifested by newborns' tendency to imitate facial expressions, is crucial in the acquisition of ToM. This ability enables the child to find a close analogy between her/him and other human agents. Rogers and Pennington (1991) have also suggested that imitation, emotion sharing, and theory of mind are increasingly complex expressions of the ability to form and co-ordinate representations of self and others.

The final candidate behavior for being the earliest manifestation of functioning ToM competence is *pretend play.* In typical development, both active pretend play and understanding of others' pretend emerges around the age of 18-24 months. As Leslie (1987) argued, that the emergence of pretend-play at this age is the manifestation of early ToM capacity, since representing someone as pretending requires similar metarepresentation as belief attribution. According to Leslie's famous example when a child sees his/her mother talking to a banana pretending that it is a phone, the child creates the further metarepresentation: My mum pretends that the banana is a phone. Leslie argues that if the child would not use the metarepresentation, it would cause that the child would mix the banana with the phone. In order to support this claim, Leslie highlights the analogy between the three types of pretend play: *object substitution* (see the banana-phone example) *pretended property*, (when the child bandages his/her doll's leg as if it was broken), and *pretended*

object, (when the child pretends that he/she drinks something from a cup) and the three basic logical properties of propositional attitudes (referential opacity, non-entailment of truth, and non-entailment of existence, respectively). However, there is a debate whether or not pretend play requires metarepresentation (e.g. Perner, 1991; Lillard, 1994).

2.2.1.2. Are these early social skills really the precursors of ToM?

In spite of the relatively rich theoretical background of the theory of mind precursors there has been surprisingly little direct empirical studies testing and supporting the idea. Most empirical evidences are indirect and found only correlational connection between language acquisition and joint attention, imitation and pretend play (Carpenter, Nagell, & Tomasello, 1998; Tomasello & Farrar,1986; Ungerer & Sigman, 1984) (see also Chapter 3). As we will see there is also evidence between language and theory of mind around the ages of 3-5 years, but certainly it does not necessarily mean that similar connection exist between the possible precursors and later theory of mind ability.

The very few empirical studies involving typically developing children, as we will see, have their own limitations too. According to my knowledge there have been only two longitudinal studies that systematically tested the potential precursors of theory of mind. The first longitudinal study (Charman et al., 2000) tested whether joint attention, imitation and play that evoked social reference measured at 20 months are the precursors of later theory of mind. It was found that only joint attention correlated with later ToM performance. However the small number of children (13) and the tests used to access joint attention (not very commonly used teasing and blocking tasks) give a limitation to the study. The other longitudinal study is very recent (Calonnesi, 2008) and tested the relation between early joint attention and later ToM capacity, more specifically whether pointing gesture is the precursors of later ToM capacity (measured with a series of tasks).

In summary, it has been found that certain, but not all aspects of joint attention correlate with later theory of mind ability, however because of the small number of studies, the small sample size of the studies and the different aspects of joint attention involved in them, further studies needed to get a finer picture about the relation between joint attention, or certain aspects of it and theory of mind.

2.2.1.3. Implicit (?) theory of mind

In the last few years there is an increased interest to measure a new behavior as a response to false belief tests instead of verbally answering questions, and this behavior is looking, measuring both first look and looking time as variables. Since these methods use false belief tests, I am going to focus on this possible early form of ToM in more detail than I did with the precursors above. The schema of the most frequently used false belief test, the so-called location change task is, that there are two separate hiding places (e.g. a box and a basket) and two characters. The first character has an object (e.g. a ball) and puts this object in one of the hiding places and leaves. While the first character is away the second character replaces the object to the other hiding place and leaves too. The question is where the first character looks for his/her object when he/she is back[2]. The method of measuring looking behavior as a response is used in infancy research in numerous different paradigms to address all kinds of research questions at this early stage of development (from as young as 1-2 month-old infants). Looking time paradigms however have been criticized in the recent years, because sometimes they are difficult to interpret (see Cohen, 2004 and I will also discuss this problem regarding false belief tests in this section), and the disadvantage of this measure can be tracked back in paradigms using false belief tests too. The idea behind using this behavior as a respond to false belief tests is that the lack of correct verbal responses found under the age of 4 may not be due to the lack of false belief understanding or more generally to the lack of theory of mind ability but rather to the way the tests are presented and the responds are required. For instance, following a verbally told story and especially verbally answering to questions may be too difficult for kids younger than 4 years and this may obscure their real abilities of understanding false beliefs. The very first study that used looking behavior, or more specifically anticipatory first look in a false belief test was Clements and Perner'study (Clements & Perner, 1994). They tested children with a modified version of the original location change task. They changed the task in two respects, first, after telling the story to the child, at the critical moment, when the first character came back, but before verbally asking questions from the child, they added a sentence as a verbal prompt: "I wonder where he *(the first character)* is going to look." And then they recorded the child's direction of first look. After that, Clements and Perner asked the children the usual questions of false belief tests and recorded the verbal answers too. The second modification was to make sure that the child

[2] More in detail about the logic and history and critics of false belief tests at the end of the Chapter 2.

remembers the story; they added three memory control questions (1) where the first character put the object (2) where the object was then and (3) whether the first character saw the object being replaced. These memory questions were asked in the middle of the story, after the replacement (before the verbal prompt) and they told the story again and again as long as the child was able to answer these memory questions correctly. Each child was tested with one false belief understanding and one true belief understanding task. The role of the true belief understanding task was to avoid that children could pass the FBT simply using some schema, (e.g. the puppet will look for the object in the empty box, no matter what) and to avoid that the looking behavior would represent some memory artifact (e.g. children would look at the place where they last saw the first character). Clements and Perner found that children as young as 3 years of age, but not 2,5 year-olds, looked correctly at the empty box after the verbal prompt, but children could correctly answer the verbal questions only from the age of 4 (see Figure 1).

The authors interpreted their results that the ability manifested in the looking behavior is the so called implicit theory of mind, versus the ability manifested in the verbal answers, which is the explicit theory of mind. Therefore they clearly differentiate these two types of knowledges/capacities, defining the implicit type of knowledge as nonverbalizable, and reject the possibility that their result, that children can 'pass' false belief tests one year earlier if a looking behavior is required, is simply due to the different level of difficulties found in the two tasks. Note, that this explanation could lead other authors to suggest that both types of measurements (looking and verbal respond) reflect the same (explicit) theory of mind. However, a more interesting question about this implicit theory of mind is the nature of this knowledge. Clements and Perner suggest a speculative explanation making a distinction between "representing a fact and making judgments about that fact", referring to the first as implicit and to the last as explicit theory of mind.

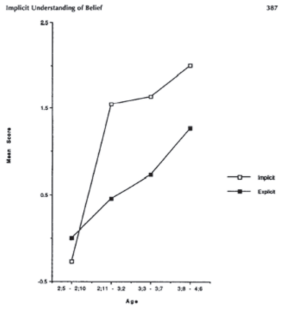


Figure 1. Mean implicit and explicit understanding scores (Figure from Clements and Perner, 1994)

A new approach of 'implicit theory of mind' research appeared in Onishi and Baillargeon's study (Onishi & Bailargeon, 2005). This study differs in many ways from Clements and Perner's results and approach. The most eye-catching (conspicuous) aspect of differences is the age difference found at passing the implicit false belief understanding test. In contrast to Clements and Perner's result (2;11 years), Onishi and Baillargeon found that infants as young as 15 months are able to pass such tests – note that the difference is more than a year and a half. This age difference, of course was at least partially due to some important methodological alternations. First of all, the entire test was completely nonverbal; the infants looked at a screen where they saw an actor performing of false and true belief situations. Another important difference in methodology was that instead of measuring anticipatory looking, the authors used the violation of expectation method. In the study infants were familiarized with an actor's goal; to get a slice of watermelon. After the familiarization trials each infant saw one of the four different belief trials: two true belief and 2 false belief trials (see on Figure 2).

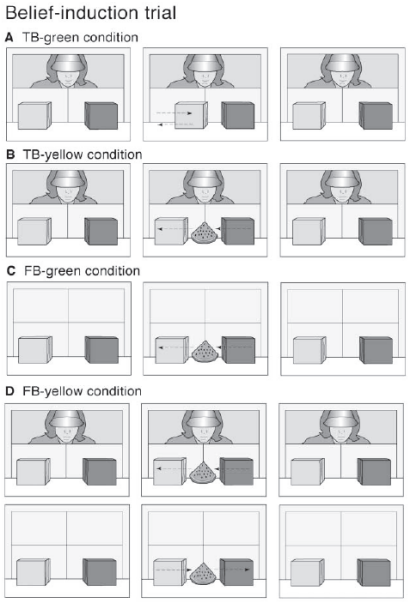


Figure 2. Events shown during the belief-induction trial in the (A) TB-green condition, (B) TB-yellow condition, (C) FB-green condition, and (D) FB-yellow condition. (Figure from Onishi & Baillargeon, 2005)

And finally, in the test trials the actor either reached into the box, which was consistent with her belief (yellow box in the TB-yellow and in the FB-yellow conditions, and green box in the TB-green and FB-green conditions) or to the other box, which was inconsistent with her belief. The actor's action, the watermelon's location and the actor's belief status were random during the trials. In all four conditions infants looked longer when the actor's behavior was inconsistent with her belief. The authors interpret their results that 15-months-old infants demonstrated a representational theory of mind "at least in a rudimentary and implicit form" (p. 257) in their study. They argue that infants at this age "realize that others act on the basis of their beliefs and that these belief are representations and may or may not mirror reality" (p. 257).

We saw that both the Clements Perner and the Onishi Baillargeon paper refers to the ability they captured, as implicit theory of mind but the age difference in the two studies is remarkable as I mentioned earlier. What is responsible for this difference? Simply

methodological issues or that the two studies represent two different abilities or different forms of the same ability?

As I mentioned earlier one methodological difference is that Onishi and Baillargeon used a completely nonverbal test to access false belief understanding. This definitely allowed younger children too to follow the story. Is it possible that the verbality of Clements and Perner's study lead to the testing of a different aspect or form of implicit theory of mind? I will leave this question open for a while and will discuss it later in part 2.1.1.4.

Another distinction in the methodologies is that Onishi and Baillargeon used the violation of expectation paradigm however Perner and Clements used preferential looking paradigm. This, itself should not be responsible for the age difference, since infants behavior can be accessed with both paradigm at a very early age – many studies used this paradigm already at 2-3-month-olds or even with newborn babies). But it is certainly possible that one paradigm is able to measure at least slightly different phenomena in a false belief test than the other. For instance the Onishi Baillargeon study has been criticized that in the testing trial what they really measured was not a specific false belief attributed to the agent but rather an attributed ignorance to the agent. However studies showed that it is counterintuitive, but children expect agents to get the answer wrong rather than chance level in an ignorance situation (e.g. (Ruffman & Keenan, 1996). This means that infants might have looked longer when the agent's behavior did not match with her (the agent's) beliefs of the toy's location not because they expected the agent to search in the other location – which would suggest false belief attribution – but because they did not expect her to search in the location where the toy really was, and therefore to get the correct answer. In other words when infants attribute ignorance they do not expect agents to search in the location where they put the object and therefore falsely believe that the object is still there, but they expect the agent not to search in the location where the object really is, because the agents should not know where the object is. The expectation of violation method is not able to distinguish between these different interpretations, and this later interpretation is more parsimonious, expects fewer skills from the infants. Other interpretation of the results is also possible, the authors themselves admit, that infants performance could be based on superficial expectations, namely, that the agent on the screen will look for the object where she last saw it disappear (also in Perner & Ruffman, 2005). This interpretation obviously does not involve beliefs about other agents.

All this does not necessarily mean that just because Clements and Perner used preferential looking paradigm their results, and especially their interpretation, should be completely

accepted. The fact that some studies found that children as young as three years of age can pass verbal false belief tests (e.g. Siegal and Beattie, 1991) suggests that the aspect of theory of mind accessed in the study is rather explicit than implicit – described by Clements and Perner above. Back to the question of whether implicit or explicit theory of mind was measured in the Clements and Perner study; if the preferential looking shows explicit theory of mind indeed, then what is the reason for the distinction between the results found with the looking behavior and the verbal response? The test was presented the exact same way in both cases; the only difference was the verbal questions asked from the children. Is it possible that this question or a specific aspect of the question itself is responsible somehow for the delay found in the verbal response? Southgate et al (Southgate, Senju, & Csibra, 2007) suggest that the word "where" is misleading for the children around the age of three, and they interpret it as where the object really is, because this is what they are used to in everyday life; the where question refers to the actual location of things. And it is not until the age of four that children can interpret the where question in the false belief tests correctly.

A recent third study tried to alloy the advantages of the above described two studies and answer the open questions regarding the theory of mind ability of infants. Southgate and her colleagues designed a study, which used preferential looking as a dependent variable to make the interpretation of the results less ambiguous than they were in the Onishi and Baillargeon's study, or the Surian, Caldi and Sperber study (Surian, Caldi, & Sperber, 2007). The Surian, Caldi and Sperber (2007) paper will not be discussed in the Dissertation since it used very similar methodology as the Onishi and Baillargeon's study (therefore the same critics are valid for this study, too). The Southgate study included two false belief conditions – watched by two different groups of children. Both conditions had two familiarization and one test phase. The two familiarization phases were identical in the two conditions. In the first one, a puppet hides a toy from two boxes into the left-hand one, while a person is watching it, then leaves. Then the person reaches into the left-hand box for the toy. The second familiarization trial is the same, except that the puppet hides the toy into the right-hand box. The first group of children then saw a false belief test trial, where the puppet puts the toy into the left-hand box, then while the person is still watching, opens the same box places the toy into the right-hand box, and closes the lid. The puppet returns to the left-hand box and closes the lid. Therefore the last location of the person's attention is the left-hand box. After that, the person turns around, so she cannot see when the puppet takes out the toy and leaves with it. Thus the baby saw that the ball was taken away but the person had a false belief that the toy is still in the right-hand box.

The second group of babies saw a false belief trial, where the puppet puts the toy into the left-hand box, then the person turns around so she can not see that the puppet first puts the toy into the right-hand box, then takes it out and leaves. Therefore the last position of the toy in the scene is the right-hand box. Thus the child saw that the ball was taken away but the person had a false belief that the toy is still in the left-hand box. In summary, in these two test trials the location where the person falsely believes that the toy is and low-level cues like the last location of the person's attention or the last location of the toy in the scene were distinct. Since the toy was always removed, the children could not look at its location, therefore it helped to avoid such tendencies from he children and also made the interpretation easier. As both locations were incorrect, children's responses could not be based on ignorance. However, Southgate et al still found that children in both groups looked at box where the person falsely believed that the toy was. The results are promising, however there were only 20 kids in the entire study, which means only 10 babies per group.

Figure 3 presents the studies focused on false belief understanding under the age of three, which sometimes also called as implicit theory of mind. Using relatively or completely different methodologies these studies found that infants or toddlers can attribute false beliefs to others at quite different ages. Since this field of research is quite new further research is needed to develop a unitary methodology and, what might be even more important, a precise terminology and description what these data reflect.

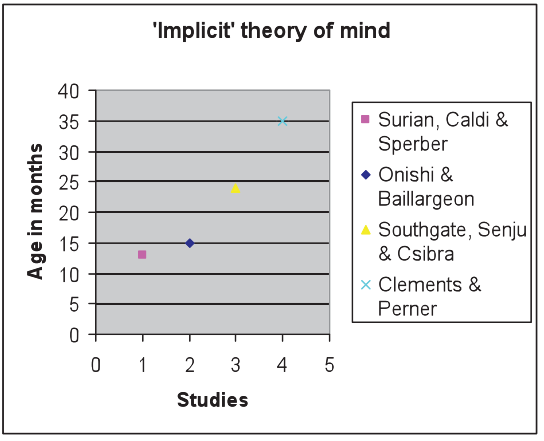


Figure 3. Studies on false belief attribution under the age of three, using eye movements as a response.

2.2.1.4.1. But what is implicit theory of mind after all?

Other than methodological issues a very important question is; what the nature of these false belief tests measured by eye-movements. Different researchers can have extremely different standpoints on this issue; at one extreme one could be stressing the difficulty of interpreting results based on eye-movements (either looking time or violation of expectation) and question the connection between this type of behavior and theory of mind ability (e.g. Katherine Nelson, 2007). One of their arguments points out that in natural every day life children do not show such tendencies of attributing mental states to others, so why would they perform it only among strict experimental circumstances? Another assumption refers to the ability manifested in these studies as "implicit theory of mind".

As I mentioned above at least two different usages of this term can be found in the literature. The one used by Clements & Perner, stresses the difference between judgmental and nonjudgmental knowledge, where nonjudgmental refers to the implicit theory of mind and judgmental to the verbal answers. According to the authors, similar results were found both in the developmental and the nondevelopmental literature, e.g. when difference was found between children's reading time and verbal report (Zabrucky & Ratner, 1986), or in the well published blindsight phenomena (Marcel, 1993). In contrast, Onishi & Baillargeon suggest that their results support the idea of an at least rudimentary and implicit representational theory of mind, however they do not explain precisely what it exactly means. They believe that this is an innate "abstract computational system that guides their interpretation of other's behavior" (pp. 257). Another possible assumption - also suggested by Milligan et al, 2007 - could be based on Karmiloff-Smith's model (1992), according to which the implicit knowledge is embedded into tasks and becomes explicit and available to conscious access and later to verbality through the "representational redescription". Further studies are needed to get closer solving this question. However, Southgate, Senju & Csibra avoid mentioning the problematic term of "implicit" when referring to 2-year-olds ability of false belief attribution, it is still not known how this ability develops later and what its relation to the later false belief understanding at 3-4 years of age; for instance whether it has a U-shaped development or a linear one.

In summary, there are still a lot of questions about the studies suggesting that already infants are able to attribute false beliefs to others. These questions are both methodological and terminological, moreover, its later development and the relation to other abilities is not studied yet. All of these should be examined and studied more in order to get a deeper understanding what these data tell us about children's social ability.

2.2.2. Belief understanding: true and false

Clearly, attribution of beliefs has a unique importance in the theory of mind literature – although the definition of theory of mind refers to a much more general concept of mental states. The reason of this can be found in the development of mental states. The earliest stage is the attribution of the *perceptual mental states*, such as to see and hear. Studying children's understanding of perception and perspective-taking is one of the earliest/oldest field in the history of theory of mind research, however nobody referred to the phenomena like this. Already Piaget (1967) has recognized the importance of attributing perspectives to others; however Piaget believed that due to children's egocentrism, it is not until the age of seven that children can attribute a perspective different from the one they have. Later, research of Flavell (1977) found that under the age of two children already know what the other person can or can not see, also called as first level perspective-taking, but it is not until the age of three or four that they know how other's perspective look like, also called as second level perspective-taking. For instance if the picture is upside down, then the other person can see it as upside down, even though the children did not see the picture that way, only in its normal position. The next stage of the development of mental state attribution focuses on *desires*. Probably the most famous study of this specific research field is the one Repacholi and Gopnik (Repacholi & Gopnik, 1997). They found that already 18-month-old babies can attribute a desire to someone that is different from their own desire. However these results could not be replicated yet (Müller, Zelazo, Frye, & Lieberman, 2002). Desires differ from perception and beliefs not only in the developmental route. Using Searle's (1983) distinction, Wellman et al. (2001) argues that in (both perception) and belief the mind is fitted to world, e.g. someone has a belief about an event in the world. In contrast, desires are about fitting the world to the mind, since desires are about making some changes in the world so as it fits to the desired event, e.g. if someone has a desire of eating a cookie, he/she has to put the cookie into his/her mouth and therefore to change the world to fulfill the desire. The final stage is therefore understanding and attributing *beliefs*. Some researchers, mainly the supporters of theory theory suggest that perception and desires – both understanding and attribution – did not simply appear earlier in development but there is causal connection among them and therefore they are prerequisites of belief understanding (e.g. Watson, Gelman, & Wellman, 1998).

Now we know why beliefs have special role in theory of mind literature but what does make false belief special? To answer this question we need to review the history of false belief

understanding test. Interestingly the story started with a study where the participants were not humans but chimpanzees. Premack and Woodruff's study in 1978 was the first, which addressed a question about the existence of theory of mind: Do chimpanzees have a theory of mind? They concluded that the evidence based on deception was sufficient to show that they do. The article generated a debate calling in not only psychologists and primatologists but also philosophers of what the minimum complexity of a task that can reliably and validly access theory of mind ability. Daniel Dennett (1978) came up with the idea that the only situation when a protagonist's belief does not concur with the reality of the world is false belief situations. It is crucial, since if they are not different it is impossible to decide whether the child's (or primate's) answer was based his/her own belief about the world or if it was based on the belief attributed to the protagonist. Later two types of false belief tests (FBT) spread in the literature, the so called location change and the false identity FBT. The *location change FBT* was first developed by Wimmer and Perner in 1983 and reached its final version in 1985 by Baron Cohen, which is also called the Sally-Ann test. The logic of the test (see also in Figure 4) is that one of two puppets places an object into one of two hiding places, while the other is watching. Then the first puppet leaves. While she is away, the other replaces the object into the other hiding place than she leaves too. At this point the first puppet comes back and the child is asked where the puppet thinks the object is (test question).

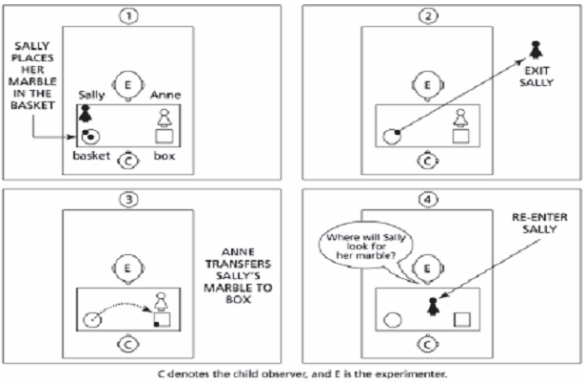


Figure 4. The Sally-Anne test from Baron Cohen et al (1985).

After the test question, usually some control questions check if the child:

- remembers where the first puppet placed the object (memory question)
- knows where the object is now (reality question)
- remembers the puppets' name (identity question)

It is also usually asked from the child why he/she thinks the object is in that certain place, called the justification question, but correctly answering the question is not a criteria of passing it. Usually children cannot answer the justification question correctly under the age of six.

The significance of the control questions becomes clear if we examine the possibility of false negative and false positive results of the test. False negative is when the child can not answer the test question not because he/she can not attribute false beliefs to others but due to some other reason e.g. did not pay attention to the story or did not remember every detail of it. A result is false positive when the child does not have the ability to attribute false belief to others but passes the test question. Note, that the possibility of passing the test question is 50%! If children cannot answer any of the control questions correctly their test cannot be evaluated regardless of passing or failing the test question. Therefore, control questions are able to filter *some* proportion of these false results. As I mentioned earlier the other type of false belief test is the false identity FBT, also known as the Smarties test. In the original test by (Perner, Leekam, & Wimmer, 1987) the child is presented a Smarties box and asked what in the box is. Then the experimenter opens the box and shows to the child that there is a pencil in the box and closes the box. Then he asks the child what another person, who was not present when the box was open, and therefore she/he has not seen what in the box is, would say was in the box. The child is also asked what he/she said when the experimenter first showed the box to him/her. A further control question checks if the child remembers what in the box really is.

Certainly the FBT went through numerous modifications during the last 25-30 years. Thank to its extreme popularity at least 100-200 studies have been published during this period (Wellman, Cross, & Watson, 2001). One of the aims of these studies was to find the simplest FBT and therefore to lower the age of the passers. A significant proportion of the studies concentrated on making the FBT linguistically simpler and unambiguous. One and maybe the most well know modification was adding the word 'first' to the test question by Siegal and

Beattie (1991). Where will Sally look for his marble, first? The authors found that this simple modification clarifies the test question for the children so they passed the FBT significantly earlier, around the age of three. However, their results sometime could not be replicated (e.g. Clements & Perner, 1994). I chose this study as an example not only because it is one the most well known modification but also because it represents the history of many other studies with similar aim. Wellman et al (2001) performed a meta-analysis to systematically investigate the two major questions of FBT literature. (1) When do children pass the FBT reliably above chance level and (2) Are there any modifications that significantly lowers this age? 178 studies with 591 conditions were included in the meta-analysis, from many countries – however the US and the UK were overrepresented, they gave 50% of the studies. The answer to the first question is presented in Figure 5 (the small circles represent the conditions of the studies and the curve represents the likelihood of 50% that a child will pass the task). We can see that children start to pass the FBT above chance level around the age of four. Around the age of three only approximately 30% of the children can pass the test. In spite of the many modifications during the past years, there is a strong consensus that children pass the FBT based on the above described logic, around the age of 4-5 (reviews: Wellman et al, 2001; Milligan et al, 2007).

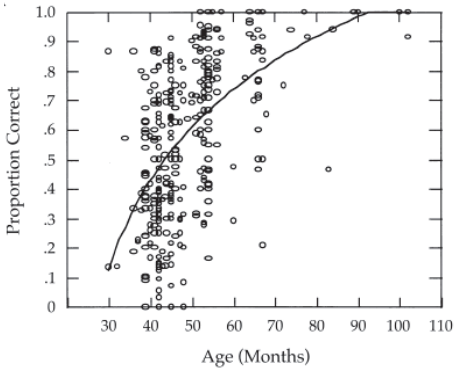


Figure 5. The patterns of data found in Wellman et al's (2001) meta-analyses. Figure from Wellman et al (2001).

We can also see on Figure 5 that the variability across conditions is large. Wellman et al made further analyses to find out whether any of the modifications make the test easier for the 3-year-olds but not for the 4 and 5-year-olds, because this would suggest that 3-year-olds

performance is suppressed by the test features, such as the child's actively participates in the task (e.g. makes the essential transformation) or the object ceases to exist (e.g. the chocolate is eaten). No such modification was found. Wellman et al interpreted their data that they support the idea that mental state understanding undergoes dramatic change around the age of 4, just like the approach of theory theory has hypothesized. This interpretation started a debate, arguing that the age shift can be due to the development of either general cognitive skills (Moses, 2001; Scholl & Leslie, 2001); or to the development of another cognitive ability; language (Astington, 2001). This phenomena can be explained by other developmental models, e.g. with modularity.

Another criticism of Wellman et al's meta-analyses is that they put an equation mark/ equal sign between the false belief test and theory of mind. This is an important criticism since many of the researchers made and make the same mistake. Obviously, theory of mind is much more complex than passing a false belief understanding test. First of all, a pure test situation in a laboratory will never be able to simulate the complexity of the social interactions of the every-day-life - and it is usually not its aim or task either. The consequences of this difference can be observed e.g. at a small proportion of people with autism who do pass the false belief test, but clearly have difficulties applying this ability in a real social interaction (more about this in Chapter 4). We also mentioned earlier, that theory of mind is not only understanding beliefs, but understanding all kinds of mental states, such as emotions, motivations and intentions therefore the FBT is far from covering the testing of theory of mind ability. It would be ideal though, if a test battery would be constructed that could measure at least the mental states mentioned above – desires, motivations and intentions (Astington, 2001).

Therefore as we saw, theory of mind is more than passing a false belief test, but at the same time, passing FBT requires not only theory of mind ability. It is undeniable that these tests require at least some general cognitive ability, executive function (Frye, Zelazo, & Palfai, 1995), working memory (Davis & Pratt, 1995) and spatial perception. However the causal relations are very difficult to find. The literature of these abilities connection to theory of mind or false belief understanding is enormous and here I am going to focus on only one further ability that is necessary to pass the standard FBT: language. Presenting FBT verbally has a couple of disadvantages. From a practical and clinical point of view it excludes some populations from testing; namely those with impaired language abilities. The other populations, who are excluded from testing are animals. I am going to talk about the first population more regarding their performance on FBT, especially about the children with

autism spectrum disorders and developmental language impairments. The other disadvantage of using verbal false belief understanding tests becomes clear when one intends to study the relation between theory of mind or as one of its tests false belief understanding and language. It is very difficult or almost impossible to get closer to answering this question without having tests that are able to measure these abilities without using/calling in the other ability. In spite of these disadvantages, relatively few nonverbal tests have been developed and the applications of these tests did not become widespread. I will investigate the possible reasons of this phenomenon after presenting the already existing nonverbal theory of mind tests[3]. Here, I am going to focus on nonverbal theory of mind tests that require acting out or in other words off-line responses.

1. **Premack and Woodruff (1978).** Interestingly, as I already mentioned, the very first test of theory of mind was nonverbal and tested the deceptive skills of chimpanzees. In the first phase four chimpanzees learned to indicate to a naïve person (trainer) which of two opaque boxes contains food. After learning that, in the second phase two new trainers participated in the test. One of them was the cooperator, who gave the food to the chimp if it indicated the correct box, and the other one was a competitor, who ate the food himself if the chimp showed him the correct hiding place of the food. The authors hypothesized that if chimpanzees have a theory of mind they will indicate the correct box to the cooperator but they will indicate the incorrect box to the competitor. It was found that two out of four chimps learned to indicate he incorrect box to the competitor.

   Criticism. As I already mentioned earlier the study got many criticisms. The main critics was that the chimpanzee needed a couple of dozens trials to learn the above described pattern of indication, which makes it possible that the chimpanzee simply learned an association between food – cooperator, and no food – competitor, which means that a more parsimonious explanation is also possible.

2. **(Call & Tomasello, 1999).** This is probably the most well known nonverbal FBT. Children were told at the beginning that they would play a hiding-finding game. Two experimenters participated in the test, a so-called hider and a communicator person. Children were also told that a sticker would be hidden into one of two containers, and the child should find it, and then she or he can keep the sticker. The basic idea of the

[3] Here I refer to theory of mind tests as tests that measure different aspects of theory of mind, not necessarily false belief understanding.

test was that the hider person places the sticker into one the containers and then at a certain point the communicator indicates the location of the sticker by placing a marker (a wooden block) on the top of the appropriate container. The test had three phases. In the first, the *pretest phase*, children could learn that the hider places the sticker into one of two containers behind the partition and the communicator indicated them where the sticker is by placing a marker on the top of the correct container. The second phase was the *control phase*, in which children were tested whether they (1) could follow the sticker as the hider replaces it from one container to the other – visible displacement condition [4] (2) could follow the sticker as it is invisibly displaced – invisible displacement condition[5] and (3) could ignore the communicator's indication when it is known to be incorrect, since the communicator did not see that the sticker has been displaced, while the children could see it – ignore communicator condition[6]. Children had to pass the control task in order to participate in the last phase. And finally in the *false belief phase* of the test children were presented with a verbal and a nonverbal FBT. The verbal FBT was identical to the ignore communicator condition, except that before the communicator returned the child was asked where the communicator would put the marker. The correct answer was certainly would put that to the wrong container, since he did not see the displacement. In the nonverbal test, however, children were presented with a test identical to the invisible displacement condition, except that the containers were switched before the communicator would mark one of them. After the communicator came back and indicated the incorrect box, the child was told to find the sticker. So the only information the child had was that the communicator must have indicated the incorrect container, since he did not see that the containers were switched – so attributed a false belief to the communicator. Call and Tomasello found that 4-year-olds were unable to pass both the verbal and nonverbal FBT above chance level, in contrast to 5-year-olds, who passed both tests well above 50%. (They also found that, apes did not pass the

[4] The hider places the sticker into one of two containers and then shows the containers to the child – so he or she could not see where the sticker was hided. The communicator then marked the correct container and left the room. While he was away the hider replaced the sticker in full view of the child. So the child had to ignore the communicator's indication and rely on his/her later experience.

[5] It was identical to the visible displacement condition, except that after the communicator left the room the instead of changing the location of the sticker, the hider simply changed the location of the two containers (so the container indicated by the communicator was still correct just had been replaced)

[6] This condition was again, identical to the visible displacement condition except that the communicator marked the container only when he came back, so after the sticker had been replaced and therefore he marked the incorrect container.

nonverbal FBT). They also found that the two FBT correlated with each other; however there was no control for age in the analyses.

Criticism. While the test was considered as nonverbal a couple of important instructions were presented verbally (e.g. information about the hiding-finding game or the find the sticker instruction). An even more important criticism is the complexity of the test, the complicated training and control trials, which could have led to the results that only 5-year-olds could pass the test and the 4-year-old failed it. Not even mentioning the difficulty of administering the test with atypical population. A further critics of the test is that it did not involve a true belief condition in the false belief phase, so it is possible that during the many trials of both the control (6 trials) and the test phase (4 trials) the child simply developed a rule based on an association: if the communicator does not the displacement of the sticker than his indication is wrong. No false belief attribution is needed to such interpretation.

3. **Gallagher et al (2000).** The purpose of this study was a little different from the others' presented here, since adults' brain activation was measured with fMRI during verbal and nonverbal mental state attribution. The authors tried to evoke the nonverbal mental state attribution by showing cartoon pictures to the participants the interpretation of which needed mental state attribution.

   Criticism. Clearly, the study had different aims from the above mentioned ones, but it verbal control questions were needed after the participants watched the cartoons to decide whether they used mental state attribution during its interpretation. Also, since the interpretation of cartoons if difficult for children, it would be very problematic to test kids in this way. Similar studies, which used cartoon stories with children, found that kids performed worse on these cartoon nonverbal FBT than on the verbal FBT (Kobayashi et al, 2008).

4. **Astington and Baird (2005).** In their study, Astington and Baird (2005) compared three different versions of the same false belief task. The first version was the standard location change FBT, in which the story was told verbally, while children were watching the puppets acting out the story. Therefore in this condition both verbal and visual information were available for the kids. In the second version, only verbal information was presented. Children could hear a narrator telling them the exact same story as in the first version, and finally in the third version children could only see the

location change, but did not hear the story, so only visual information was given to them. The authors did not find any difference in children's performance among the three versions; 84% of the children passed or failed all three false belief tasks. In accordance with the literature, most 3-year-olds failed all three tests, but most 4-year-olds passed them. However, it is not very surprising, hence in all there versions the test question and the control questions were presented verbally, and children answered these questions verbally too. It looks like that even if there were some differences how children are represented with the tasks in the beginning due to the different presentation modes, when they had to answer the verbal questions, these differences dissolved/disappeared.

2.2.2.1. True belief understanding

Because of Dennett's famous argument why false belief understanding is the minimum requirement of testing theory of mind ability, true belief understanding has never really been in the focus of the theory mind literature. However, the easiness of true belief tests, or more precisely when children can pass such a test has been a debate from early on. The debate certainly has consequences on the approaches of how the mind and therefore theory of mind develops. Some studies (Lohmann, Carpenter, & Call, 2005; Wimmer & Perner, 1983) found that children can pass true belief understanding tests between the ages of 3-4, thus earlier than the FBT, however other studies found that children can not pass true belief tests earlier than FBT (e.g. Roth and Leslie, 1998; Riggs & Simpson, 2005). The difference has - at least in part - methodological reasons since true belief understanding can be measured in different ways. In the so called "no switch condition" (or location change) after the first character leaves the scene the second character takes out the object from the container and puts it back into the same one without replacing it. This condition is usually easier for children than the other true belief conditions, since only one container is involved in the situation (however, there is another container in the scene but nothing happens to it), and therefore no inhibitory control is needed to pass it. In the "switch condition" (or location change) as its name also refers to it, the second character does replace the object from one container to the other, but the first character watches this displacement. According to Lohmann et al (2005) children around the ages of 3.5-4 can pass this type of true belief test above chance level but not significantly. A very similar logic was used in a verbal prompting condition, except that the

first character did not watch the displacement, but it was told her/him by the second character. Children found this task easier though, and passed it significantly above chance level at the same age. A third type of true belief task was used by Roth and Leslie (1998), in the "partial true belief" task, in which a true and a false belief task were combined. The first character placed an object into one of three hiding places while the second character was watching. Then the second character left, and the first character asked the child to hide another object somewhere else (to one of the two other hiding places) and than asked the child where the second character will look for the object (in reality there are two coins). The results were somewhat surprising; children between the ages of 3;0 – 3;8 performed better on the partial TBT than on the FBT but this relation has changed as children got closer to their fourth birthday 3;9-4;0 years, however the difference is not significant. Because of the low performance on both tasks, it is very probable though that, children found these tasks too difficult in general. The three hiding places and the two objects (and both were the same; coins) heavily loaded the working memory capacity and that led to the low performance.

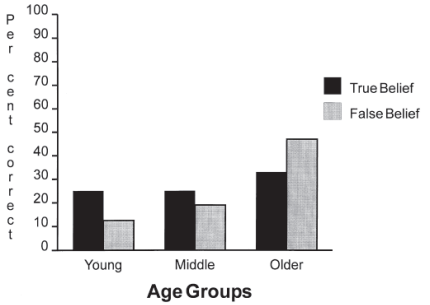


Figure 6. Percentages of 3-year-olds in the three age groups passing the 'prediction' question in the partial true-belief and false-belief tasks (From Roth and Leslie, 1998).

In line with this argument, Riggs and Simpson (2005) also found that another type of true belief task is not easier for 3-4 year-olds than the standard false belief task (children passed them at the mean percentage of 38% and 33%, respectively). After presenting a standard location change false belief test, they either asked the usual false belief question (Where does the first character think her object is?) or a true belief question, where the character's true belief was not identical to the state of reality: When the first character left the room where did she think her object was? It is obvious that the true belief question refers to the past, and it is possible that this made the TBT more difficult, but it does not explain the other result that the

authors found; significantly more children passed the memory question. We saw that there is some data, which suggest that when a true belief understanding task is similar to the false belief task in its structure and logic, then the TBT is no longer easier than the FBT. Another area of TBT application is more methodological. TBTs are sometime used as control conditions of FBTs. As we already saw in the "implicit theory of mind" part, especially if a nonverbal test is used to measure FB understanding, it can be crucial to use TBT to control that children do not pass FBT based on a simple nonmentalistic rule, e.g. the character goes to the empty hiding place or to the hiding place the character was associated with earlier (see e.g. Heyes, 1998). Nevertheless, such nonmentalistic rule can be used to pass verbal FBTs too, since the control questions do not refer to the mechanism how the child passed the task. Even if a justification question is asked (why the first character will look for her object at a certain place), it is almost never used as a control question. No wonder, since just because a child is not able to give a proper verbal description of his/her mental processes, it does not necessarily mean that these processes do not exist. Children are able to give proper answers of the justification question only around the age of 7 (e.g. Clements and Perner, 1994). In spite of this, most verbal FBTs do not use TBTs as a control task.

While this Dissertation focuses on the kindergarten age, which is undeniably a crucial period in the development of theory of mind, there are a lot of important changes after this age that are often forgotten and receive much less attention. I do not intend to indicate that around the age of 4 children acquire the ability of theory of mind and its development is finished at that point. Not only second order mental state attribution (Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994) develops later – around the age of 6 – but also the interpretative understanding of mind; that is although two people saw the same event they might arrive at different conclusions (Chandler and Carpendale, 1998), or understanding of humor and irony (Happé, 1994; Győri, 2000) and these are still just a couple of examples to picture the complexity of theory of mind ability.

Chapter 3. Possible developmental relations of language and theory of mind

In order to understand the possible relations between language and false belief understanding, it is inevitable to investigate the question from a broader picture and see what the relation between language and theory of mind can be. But before focusing on this developmental relation, I will specify what I mean under language, because as we will see, language, just like the concept of theory of mind, is a broad term and different authors may use it in different ways. First of all, it is crucial to differentiate the functions of language from the structures of language. In the Dissertation I am going to use the concept of language as a structure, in a Chomskian way, and therefore I am not going to discuss pragmatics. From a more methodological point of view, the measurement of pragmatics always require by definition the active participation of another person, which makes its operationalization more problematic and less objective.

Research regarding the relation between the development of false belief understanding and language has been in the focus of the developmental psychology at least in the last decade. The interest in this research area came from many directions. (1) architecture of the developing mind (2) developmental psychopathology (3) ethology, primateology etc. The two main questions that (almost) all of these areas are interested in refer to the nature and the direction of this connection, however as we will see, these questions are very closely related. The nature of the relation theoretically can be imagined in three ways. The first possibility is that although language and theory of mind affect each other during development e.g. they facilitate each other, none of them has a special role in the development of the other. In other words, it means that one can be acquired without the other. The second possibility is that this relation is essential; e.g. language or one of its aspects is necessary to acquire theory of mind ability. The third option is only a theoretical one, since no approach or data suggest that there is no relation between the two at all. These options, however pose further questions, which lead us to the direction of this relation. More specifically, if none of the two abilities play a special role in the developmental of the other, than what is the reason of the many correlations found between the language and theory of mind abilities. A possible explanation could be that theory of mind ability is innately specified but this competence is not evident due to performance limitations. Language is needed – among other abilities – for theory of mind to the manifestation of this competence (e.g. Roth and Leslie, 1998). This assumption strongly

implicates a modular approach of the mind, but not exclusively. A third, domain general ability can also be in the background of these concurrences, and its development manifests in the two abilities too. The possible candidates are working memory, executive function or recursion. But since almost all of the data we have on theory of mind understanding used some kind of verbal theory of mind test (e.g. FBT) we can not exclude the possibility that this relation is only a byproduct of the tests we use, but this assumption, again is not only supported by modularity theorists e.g Chandler, Fritz, Hala (1989) or German and Bloom (2000). The no special role assumption can also be explained by the theory theory; language is simply a way of providing the necessary information to construct a theory of mind (Gopnik and Wellman, 1994; Perner, 2000). On the other hand the fundamental relation is clearly more associated with non-modular approaches, since the ability of one is essentially necessary to acquire another, we can no longer talk about independent modules. But still this assumption is far from being unified. Different researchers are divided both in the direction; which ability develops first, and affects the other, and in the concerned aspect(s) of the abilities/ aspects in question.

3.1. The direction of the relation

Regarding the direction of the relation, the main question is whether either language (or any aspect of language) or theory of mind (mostly rather just false belief understanding) is a prerequisite of the other ability, which means a clear causal direction between them. However there are at least two alternations from this approach, (1) the relation is not causal and (2) the relation is bidirectional; both language and ToM affect each other during development. However, as we will see this bidirectional approach has two manifestations, a "weak" bidirectional approach, which sees the bidirectional only during development. That is it acknowledges the role of social cognition in e.g. word acquisition, but believes that during kindergarten age it is language or a specific aspect of it, has a special casual role in later theory of mind development (usually means false belief understanding) (e.g. de Villiers, 2007). Therefore the bidirectional of the relation appears at two different points of development. The "strong" bidirectional approach, however, suggests that this bidirectional relation is also present at one certain point of development, namely at kindergarten age (e.g. (Slade & Ruffman, 2005 or Perner et al, 2005).

3.2. The nature and the direction of the relation

As we saw above, when the relations of two developing abilities are put under scope, it is crucial not to focus on only one age group, but to see the process of the development of these abilities. Around the age of 1 and 2 there is a significant advance in the development of both abilities; emergence of joint attention behaviors and emergence of first words. As we already seen in Chapter 2, there is some data that joint attention behaviors are the early forms or precursors of theory of mind. It has been argued that joint attention behavior has an important role in word acquisition. Since Quine's paradox about the indeterminacy of reference with respect to the large issues of meaning it has been a debate how children are still able to learn words. Among other supporting mechanisms (e.g. using certain constraints; Markman, 1994) joint attention behaviors, such as shared eye gaze and pointing, help children to restrict or even to determine the reference(s) (Bloom, 1999).

There are many more arguments about the direction and the affected aspects of the two abilities at a later stage in their development, around the age of four. As mentioned, this is the age when children reliably pass false belief understanding tests, therefore studying the relation between the two abilities at this age has been in the focus of research in the last decade.
The usual method of these studies is that either longitudinally or cross-sectional children are tested with a couple of language tests, which usually test different aspects of language ability and theory of mind tests, which is almost always a verbal false belief understanding test and the performances on the tests are correlated or the data are analyzed with some kind of regression analyses. Regression analyses are used in the hope of finding a causal connection between the two abilities. However, as we will see the results of these studies can be contradictory, not only in respect the direction but also in the effected aspects.

3.2.1. Pragmatics

Pragmatics is the ability "to use and interpret language appropriately in social situations, which depends on keeping track of listeners' and speakers' beliefs and intentions. Thus, pragmatics and theory of mind are related by definition" (Astington & Jenkins, 1999 p.2), therefore when I talk about language I use the concept in a Chomskian way, meaning that only the formal aspects of language are included and pragmatics does not.

3.2.2. Semantics

We already saw above that joint attention plays an important role in the development of semantics, namely word acquisition during infancy, but whether later semantics impacts theory of mind development has been debated.

One line of approach argues that semantics is crucial for children in participating in verbal social interactions, which enables theory of mind development (Dunn & Brophy, 2005) or according to Katherine Nelson's terminology, their entering the "community of mind" (Nelson, 2005; 2007). However, in this context it is very difficult to differentiate between semantics and pragmatics.

Probably the most obvious connection between semantics and ToM is the development of mental state words, around the age of 2 the perceptional than the emotional and finally the epistemic mental state words (Bartsch & Wellman, 1989). For instance Olson (1988) argues that parent's usage of mental terms (or adult in general) focuses children's attention to these mental, nonobservable entities and since the same linguistic terms are applied to their (children's) own mental states and other's mental states children can map these subjective states to other's behaviors and experiences and therefore to develop the comprehension of mental states, both their owns and others'. However this mapping process has been in question/queried by others (Gopnik, 1993) but there are some data that support this type of connection between mental state words – both comprehension and production – and ToM – measure by false belief test (Dunn, Brown, & Beardsall, 1991; Ruffman, Slade, & Crowe, 2002).

Not surprisingly, the effect of general vocabulary has also been reported. A couple of studies found correlation, both in longitudinal and in cross-sectional studies between general vocabulary – usually measured by PPVT or BPVS – and (later) FBT performance (e.g Astington and Jenkins, 1999; Slade and Ruffman, 2005). However after controlling the variance of the different tasks, it turned out to be a better predictor of FBT performance than vocabulary alone.

The fourth line of approaches focuses on importance of labeling and its connection to the acquisition of ToM. An interesting study found a surprising connection between false belief understanding and word learning. (Happé & Loth, 2002) developed a so called word learning false belief task (WFBT). In this task a word learning situation is embedded in the standard FBT. The first character places a novel object into a box without labeling it, than she leaves. While the first character is away the second character replaces the novel object with his own novel object (still no labeling). When the first character comes back she labels the object

without opening the box. In order to find the correct reference of the new label, e.g wug, the child has to attribute a false belief to the first character that is she falsely believes that the object in the box is still the one that she put there. Happé and Loth (2002) found somewhat counter intuitively – more complex task is easier – that significantly more children passed the WFBT compared to the standard FBT. They concluded that the two types of FBT (word learning and standard) capture the different developmental trajectories of two distinct theory of mind mechanisms. The WFBT shows the ToM mechanism for communication and the standard FBT mirrors the ToM mechanism for interpreting and predicting behaviors. However, a more parsimonious explanation is also possible, without suggesting two ToM mechanisms, namely that labeling a novel object directs children's attention toward the relevant information, and therefore word learning situations have a facilitative effect in FBT, or maybe even in other related tests. Similar labeling effect was found in other studies (e.g Jacques and Zelazo, 2005) too.

3.2.3. Syntax

Regarding syntactic ability, the main question that encouraged the conduction of many studies is whether syntax in general or a specific aspect of syntax plays a causal role in ToM development. Main question certainly among those who believe that syntax play a crucial role in FB understanding and therefore in ToM development. The reason why syntax was attributed a special role in this debate is because it requires following and understanding how the arrangement or combination of words affects the meaning of the sentence. Astington and Jenkins (1999) argue that this syntactic ability of keeping track of, and correctly representing, often quite complex relations between individual elements of a sentence is similar to the type of ability that would help a child keep track of and represent the complex relations in a false belief task (different locations of the object, presence and absence of the characters). In support of this hypothesis they found that early syntactic ability measured by an item of TELD predicted later false belief performance even when age and other aspects of language e.g. semantics were controlled for, but the reverse relation does not hold. This was the first paper that systematically tested the question in a longitudinal study.

A more specific aspect got into the focus of research with Jill de Villiers' *sentential complement* hypothesis. Sentential complements are a type of embedded sentences where the complement-taker verb can be a communication or mental verb. The embedded part of the sentence (subordinate clause) is the complement. E.g. John said that he did not eat the cake. Children between the age of 3 and 4 years master sentential complements.

The uniqueness of sentential complements according to de Villiers is that their structure is very similar to false belief understanding. That is, they "uniquely allow the representation of false propositions" (de Villiers & Pyers, 2002), p.90) the entire sentence can be true, even though the embedded part is false. It has been a question, why the mental verbs, which refer to communication or to thinking / belief are special? Why the complement hypothesis does not hold for verbs of desires, which actually develop earlier (Bartsch and Wellman, 1989) According to de Villiers it is because they do not have the same grammatical structure (that + finitive). E.g. I want John to eat the cake and not I want that John eat the cake. After a couple of cross-sectional studies which found correlation between sentential complements and false belief understanding (de Villiers and Pyers, 1997; Tager-Flusberg, 1997) de Villiers strengthened her results in a longitudinal study (de Villiers and Pyers, 2002). The authors found that early sentential complements were the best predictor of later false belief understanding when language tests and false belief understanding tests were administered 4 times in a row with children between the ages of 3-5. However the reverse relation was not true. De Villiers concluded that the mastery of sentential complements is a prerequisite of the emergence of false belief understanding and therefore theory of mind. Later, however the theory have been refined and became more precise, for instance in that FBT is not equal to ToM and that the relation was restricted to the kindergarten ages, therefore it acknowledges the possible reverse relation mentioned above between joint attention and language acquisition (de Villiers, 2007). But the main statement, which is a very strong one, about the casual relation has not changed yet radically. The Complement hypothesis has induced relatively many studies both in English and in other languages. A methodological outlook is necessary here, since the way of measuring sentential complements is not obvious. De Villiers complements test has pairs of pictures, which are presented with short stories and test questions.

The Memory for complements task:

The experimenter reads the story pointing to the important objects in each picture and pointed again to the first picture when asking the question.

"She told the girl there was a bug in her hair, but it was only a leaf."

Q1: What did she tell the girl?

Q2: What did she tell the girl was in her hair?


Figure 7. Pictures used in de Villiers complements task (2002).

The study used elicited (verbal) imitation as a response in other words the child was expect to repeat the relevant part of the sentence that she/he just heard. The argument with this type of method goes back to Slobin and Welsh 1973 (also Bloom et al, 1974; Gerken et al, 1990). They argued that the child is not able to produce a certain grammatical structure as long as he/she has not acquired that grammatical structure. In spite of the disadvantages of the method –e.g. it is not exactly clear how precisely its represents the child's own grammar, either if he/she does imitate the same sentence or if he/she does not – the method has been wide spread as a psycholinguistic method.

3.2.3.1. Complements versus relative clauses

One of the criticisms of de Villiers' hypothesis is that it is not only sentential complements that are the prerequisites of the mastery of FBT but in more general, relative clauses. Relative clauses are subordinate clauses that modify a noun. (e.g. The man chases a cat that is small.) The assumption was tested by a couple a researchers (de Villiers and Pyers, 2002; Hale & Tager-Flusberg, 2003) but none of them found a correlation between relative clauses and later false belief understanding. However, another study, which focused on specific relative clauses, did. (Smith, Apperly, & White, 2003) argued that double-event relative clauses, such as *The girl kicked the man that jumped over the wall*, have special role in false belief understanding since "they require the handling of metarepresentation because they embed a relative clause event inside a matrix clause event" (p. 1716). But unlike sentential complements they do not require the handling of misrepresentation. But the study used a different method to test relative clause comprehension, namely truth value judgment tasks, where children had to judge whether an event acted out by an experimenter matched a spoken relative clause sentence. We already saw how expecting a judgment from kids might change their performance in FBT so it is possible that something similar has happened here too. But even more importantly, the children were not tested with the sentential complement task, so it

is impossible to tell whether that task's predictive effect would have overwritten the predictive power of the double-event relative clauses.

3.2.3.2. Studies in other languages

Studies with sentential complements were also conducted in languages other than English. The importance of these studies is to test the universality of the relation found between complements and FBT. Perner and his colleagues (2002, 2005) for example draw attention that in German language the verb want can also have the same that + definitive grammatical structure as e.g. say and think have, and children acquire the complements with this verb earlier than the complements with either say or think. Yet German children also pass FBT around the age of four and not earlier. And also the complements with the verb want does not show the same relation that the ones with say or think. Similarly, the specificity of Chinese Mandarin and Cantonese languages are different from English but actually show the opposite pattern than German, since in these languages it is possible to use the same grammatical structure to talk about beliefs and desires without the complex finite that+complement construction. Still, children start to talk about desires much earlier and more frequently than about beliefs, just like their English mates (Cheung, et al., 2004; Tardif, So, & Kaciroti, 2007; Tardif & Wellman, 2000). Perner argues that these results challenge de Villiers' hypothesis and language determinism, and suggest that while language definitely gives input to theory of mind development, there is no casual relation between the two. However, de Villiers (2005) argues that it is not the syntactic form that matters, but that verbs of desire (both [want + infinitive] in English or [want-that] in German) or pretend (e.g. [pretend-that] in English) take irrealis complements, which means that they refer to future or imaginary events. Therefore they cannot be true or false, like complements of verbs of communication and belief (realis), they can only be fulfilled or unfulfilled. According to de Villiers this is a crucial difference in acquisition, since it will cause different developmental trajectories and hence a difference in their relation to FB understanding.

3.2.3.3. Training studies

Another way to test de Villiers' complement hypothesis is to use training methodology. It is expected that if there is a causal relation as de Villiers suggests than training and therefore developing on the syntax of complements will promote false belief understanding, but the reverse relation should not hold. The first such training study was conducted by Hale and Tager-Flusberg (2003). They tested 3 groups of children between the ages of 3 and 5. When

pre-testing these children all of the groups failed on all the 3 tasks, which were: FBT, sentential complement test, and relative clause test. Children were trained during 2 sessions in a week with one of the 3 tasks (that was the bases of the classification into the 3 groups) and received feedback. After the training their performance certainly significantly improved in the trained task, the question was whether it promoted the performance in the other tasks too. The only task that had such an effect was the sentential complement task, so when children were trained on this task they not only improved significantly on this task but also on theory of mind tasks (FBT and appearance reality tasks). Importantly, the training on FBT did not have such an effect on the sentential complement task. These findings seem to support de Villiers' hypothesis, since the improvement on the sentential complement task resulted an improvement in the theory of mind tasks, but the reverse effect was not found. At the same time, the results also challenge the necessity of the mastery of sentential complements to pass theory of mind tests, since the group of children who received FBT during training showed equivalent developmental changes in theory of mind as the sentential complements training group, without having a mastery of sentential complements. Similar results were found in (Lohmann & Tomasello, 2003) training study, they also found that training on sentential complements promoted false belief understanding. And while they found that this effect was independent from the effect of conversation about deceptive objects, it is still not clear what aspect of complement syntax is responsible for this relation (e.g mental verbs) or whether the reason of this relation is the verbality of both tests, that is verbally we express both false belief and sentential complements very similarly, but metarepresentational theory of mind is not necessarily verbal.

3.2.3.4. Criticism of de Villiers' complement hypothesis

As we saw there are a couple of criticisms against the complement hypothesis, below is a summary of these arguments.

1. The difficulty of differentiation between the role of sentential complements and mental verbs
2. It is relative clause and not sentential complements that predict later FBT performance.
3. Studies in other languages (German, Chinese Mandarin and Cantonese)

4. Earlier manifestation of the mastery of FBT (around the age of 1.5-2 years)

5. Methodological issues: general language ability or grammar play a specific role in ToM development

3.2.4. General language ability

And finally, we should also consider the possibility that no language aspect has a more special relation or role in theory of mind development than another, but language as a whole, including both semantics and syntax (and in a different approach of language ability pragmatics too). Nevertheless, this approach is not unitary either, as we will see.

One of these studies directly tested though complement hypothesis. Slade and Ruffman (2005) argued that the methodology used by de Villiers (e.g. de Villiers and Pyers, 2002) was not entirely appropriate since instead of using standardized tests of different aspects of language, they used spontaneous speech (IPSYN) to access children's semantic, syntactic and general language abilities. They found that it was general language ability that predicted later FBT performance (similar findings were also found by Cheung et al, 2004), but even more interestingly the reverse direction was also found; FBT performance has also predicted later language ability. The authors conclude that there is no special relation between ToM and language; in typical development they certainly effect each others development but none of them is a prerequisite of the other, this relation is not causal. Another important finding of the paper was that neither language nor ToM test indexes showed a correlation with early working memory performance. Based on this study it seems that the hypothesis that working memory as a third factor, which would be responsible for the relation found between language and theory of mind is not supported by empirical data.

A somewhat different approach is suggested by Tomasello, he suggests, that during infancy it is social cognition that plays an important causal role in language acquisition (e.g. (Tomasello, 2000; Lohmann et al, 2005), not only in word acquisition but also in the acquisition of grammar, suggesting a usage-based language acquisition theory (Tomasello, 2003). The author suggests though, that around kindergarten age it is language or more precisely linguistic conventions in communicative interactions (discourse) that helps children to the understanding of beliefs or even false beliefs. Clearly, the theory emphasizes the communication through language and the social interaction with language, and not the formal aspects of language, or language as an abstract representational medium, the later of which is in the focus of the current Dissertation.

Similarly, according to Katherine Nelson (2005; 2007) it is discourse that enables children to become a member of "the community of mind", in other words to know they ways around and to successfully participate in the social world – a part of which is certainly false belief understanding.

ToM	Direction of relation	Author(s)	Aspects of language
Precursors of theory of mind	→	• Baldwin & Moses, 1989	Semantics
Theory of mind	←	• Dunn et al, 1991 • Furrow et al, 1990 • Ruffman et al, 2002	Semantics
	←	• Astington and Jenkins, 1999	Syntax
	←	• de Villiers & Pyers, 2002 • Tager-Flusberg, 2000	Special aspect of syntax: Sentential complements
	↔	• Slade and Ruffman, 2005 (no causal relation) • Perner et al (2002, 2005) (no causal) • Sperber & Wilson, and Happé (special relation)	General language ability
Community of minds	←	• Nelson (discourse)	General language ability
Social understanding	→**(infancy)** ←**(kindergarten)**	• Tomasello (discourse)	General language ability

Table 1. Possible developmental relations between language or certain aspect(s) of language and social cognition, or certain aspect(s) of social cognition.

Chapter 4. Relevant neurocognitive developmental disorders

How will the connection between language and theory of mind development alters/differs from the one we found in typical development change if one of these abilities has an atypical developmental trajectory? What developmental patterns can be found in children who have deficit in one of these two abilities? The two relevant neurocognitive developmental disorders regarding language and theory of mind are autism spectrum disorders (ASD) and developmental language impairments (DLI or more specifically, specific language impairment (SLI)). Since Baron-Cohen's famous paper: "Doest the autistic child have a theory of mind? " in 1985, it became a well-known and accepted phenomena that most people with autism have a deficit in theory of mind. In sum, ASD and DLI were chosen as candidates of the selective impairments of the two developmental abilities in question, theory of mind and language.[7] Since the Dissertation mainly focuses on typical development, the atypical populations will be present only in 2 of studies, I will only discuss the most relevant literature from the Dissertation's point of view of these two developmental disorders.

4.1. Reasons to study neurocognitive developmental disorders

Neurocognitive developmental disorders are a type of developmental disorders where the disorder has neural background, cause and it has impact on the developing cognitive system, creating behavioral symptoms. Clearly, a bigger attention has been drawn to neurocognitive developmental disorders, such as autism spectrum disorders, ADHD, Williams syndrome, language disorders, Down syndrome etc. in the last 2-3 decades. As an interesting result, only in year of 2007 more than 1000 papers were published in an autism related topic. One of the motivation is certainly to better understand these disorders, which is crucial for effective treatment and intervention. The growing interest in the better understanding of these disorders is largely due to the growing number of children diagnosed with these disorders in the last decade (see autism as an example).[8] On the other hand, another motivation is the better understanding of the typically developing mind. The idea is that the developing mind in certain syndromes has a selective deficit or impairment. Studying this selective deficit and the other, supposedly intact cognitive mechanisms – called the residual normality hypothesis –

[7] Different approaches regarding the inferences of neuro-developmental disorders to the typical development will be discussed later.

[8] Here I am not going to details of the reason of this increasing prevalence, but it is probably due to both the better diagnostic equipments, trainings etc. and the increasing number of cases

might help us to better understand the specific mechanism and its connection to other cognitive systems, which can apply to normal development too. This approach is closely associated with the modularity theory. A major argument of modularity is developmental double dissociations. Double dissociations, "the chocolate cake of neuropsychologists" (Karmiloff-Smith, Scerif, & Ansari, 2003) have been studied in neuropsychology to improve the independence of two systems. In these rare, even single cases an already acquired system has been damaged, leaving the rest of the cognitive system intact. If two patients are found, one with damage in one of the systems only and the other patient is damaged in the other system only, then one can argue that the systems are independent from each other (see e.g.). The same line of arguments has been used in developmental double dissociations, too. For instance, Williams syndrome and developmental prosopagnosia have as a double dissociation implies the independence of face processing structures from general visuo-spatial processing (Pinker, 1999). Similarly, it has been argued in the 1990s, that double dissociations can be found between the formal and functional aspects of language. In has been argued that in autism the formal aspects, such as syntax, phonology or semantics is, at least in some cases, intact, while pragmatics is, by definition, impaired. Specific language impairments have been argued to show the exact opposite pattern of deficits, intact functional aspects, impaired formal aspects of language. Even more specifically regarding the topic of the Dissertation it has been argued that the double dissociation found between autism and SLI regarding language and theory of mind suggests the independence of two abilities. There has been substantial criticism however, which question not only the line of the arguments, or several aspects of the arguments, but also the existence of developmental double dissociations.

4.2. Autism spectrum disorders

ASD are pervasive neurodevelopmental disorders. The latest prevalence of autism is 1 in every 150 individuals and affecting four times as many males than females. The diagnosis of autism is based on behavioral signs, and it is characterized by the so-called Wing's triad (Wing & Gould, 1979): impaired social interaction and communication and restricted repetitive behavior. The diagnostic criteria of autism are that the behavioral symptoms have to be present before the age of three. Although these behavioral impairments are present in every child, autism is a very heterogeneous disorder in terms of the severity of the symptoms, the general cognitive abilities and the collateral deficits. Therefore in research it is worth to differentiate at least 3 subgroups of autism.

- Asperger syndrome (AS): it is named after Hans Asperger, an Austrian pediatrician who in 1944 described four of his children patience with the syndrome (however there is still some contradiction in the literature what was exactly described by him). It is still controversial, whether AS is a discrete subtype of autism or if it is continuous with other forms of the condition (Volkmar & Klin; 2000). Within the ASD the syndrome is associated with non-impaired general intelligence, non-impaired formal language (but impaired figurative language, pragmatics), and higher level of adaptive and self-supporting skills. It is differentiated from high functioning autism with the lack of language delay.
- High functioning autism (HFA): as mentioned above AS and HFA are very similar disorders, except that people with HFA do have a language delay.
- Low functioning autism (LFA): people with LFA have impaired general intelligence (IQ under 70), impaired formal language and the probability of other collateral symptoms is higher.

In line with this heterogeneity the background reason of autism can also vary. ASD have dominantly genetic origin, however little is known about the specific genes that are affected, whether it is caused more by multigene interactions or by rare genetic mutations or by both. Environmental factors, such as viral infections and exposure to environmental chemicals were also associated with autism, just like problems during pregnancy or delivery, but again, no universal and autism-specific "trigger" was found.
Because of the behavior-based diagnosis, cognitive-behavioral research has always had and has a unique role/importance in autism research.

4.2.1. Theory of mind deficit in autism

The theory of mind literature regarding autism is enormous. The Dissertation raises very specific questions about this population, namely how language and FB understanding connects during development. Since the discussion of the entire topic of 'theory of mind in autism', which includes the cognitive models of autism for instance, is beyond the scope of the Dissertation, I will strictly concentrate on the most relevant literature of the topic.

As mentioned above, the idea of theory of mind deficit in autism was first published in the influential paper of Baron-Cohen, Leslie and Frith (1985). Briefly, the study found that the vast majority of children with autism (16 out of 20), but not children with Down syndrome or typically developing children, failed on the first order false belief test, on the Sally Ann test (remember that failing on the FBT means that they passed all control questions). The authors concluded that children with autism "fail to employ a theory of mind" (Baron-Cohen at al, 1985, p. 7). The paper went through a lot of criticisms since it has been published, mainly due to the wide age range in both the autism (e.g. CA: 6;1-16;6 years) and Down syndrome (CA: 6;3-17;0 years) group, and the heterogeneity of the these groups. Also nowadays Down syndrome is not a preferred control group of autism, because of the uneven cognitive profile that was found in the syndrome. In spite of these criticisms, the idea that people with autism have a deficit in theory of mind is still a valid statement, however the picture of theory of mind in autism has become much more sophisticated.

One of the most important results of the theory of mind ability in autism since 1985 was Happé 's finding in1995 and in an even more precise study from 2005 (Fisher, Happé, Dunn, 2005). Her study intended to shed light on a phenomena found already in the 1985 paper, namely that some children with autism do pass false belief tests. The so-called "problem of the passers", where the problem refers to the question, that if children with autism universally have a theory of mind deficit, how they can pass false belief understanding tests? One of the explanations suggested by Happé (1995) is that they use some kind of verbal compensatory strategy. She found that children with autism tend to pass FBT above the verbal mental age of 7 and she found a ceiling effect at the verbal mental age of 13. However, the exact mechanism of this verbal compensatory strategy is still unknown, but the strong correlation between verbal ability and false belief performance in autism has been repeated since (Fisher et al., 2005).

They have the ability but it is very fragile, so under pure test circumstances they can pass these tasks, but under "noisy" circumstances of real life they can easily get confused (noisiness: many variable in the environment, plus more people, bigger pressure, anxiety, but this is very speculative).

So as we have seen above language and the performance on FBT are highly associated in autism, the possibility of a verbal compensatory strategy was also suggested (Happé, 1995, Fisher et al, 2005). Thus studies with a nonverbal FBT are crucial to have a better insight to the theory of mind ability in children and people with autism. According to my knowledge there has been only one paper that conducted such a study. Colle, Baron-Cohen and Hill

(2007) applied Call and Tomasello's (1999) nonverbal FBT, however in addition they used 3 true belief and 3 control tasks as well, to investigate the false belief understanding of 3 groups of children: children with LFA, with SLI, both groups of children's language production and verbal comprehension were maximum equivalent to two years of age, in other words these groups were matched to language ability, and they were also matched to their chronological ages (8.1 and 8.3 years). The third group of children was typically developing children, whose nonverbal MA was matched to the autism group's nonverbal MA (4.9 years). The results were not surprising, the only group difference found was with the FBT, and post-hoc analyses showed that it was due to the autism group's low performance. As I just mentioned the results are not surprising. Children with autism who pass FBT have relatively good verbal skills (Happé, 1995; Fisher et al, 2005), so it is no wonder that children with the verbal MA of two, did not pass it. The SLI group's nonverbal MA (and chronological age as well, due to the diagnostic criteria) was very high, more than 8 years, so again it was no surprise that hey passed the FBT. Therefore the test did not answer two important questions (1) whether children with ASD with the verbal MA of minimum of 7 years would pass the nonverbal FBT, which would rise questions on the verbal compensatory strategy and (2) children with SLI would still show a delay in FB understanding even with a nonverbal FBT, which would suggest close relation between their language ability and theory of mind ability.

An even more recent study however tested more relevant questions from the Dissertation point of view. (Senju, in press) using the eye tracking method developed by Southgate and colleagues (2007) (also discussed in Chapter 2) to test if 6-to 8-year-old children with ASD understand false belief if it is presented nonverbally and the measured response is spontaneous anticipatory looking. They found significant difference between the typically developing group and the ASD group in both the verbal and the non-verbal eye-tracking FBT, but when verbal mental age was covaried out of the analysis, the different was not significant. The authors argue that just like in Happé's study (1995), children fail to pass verbal FBT under the verbal mental age of seven (the mean VMA in the ASD group was 6;3), but based on the data of the verbal FBT it is unclear whether it is only due to the verbality of the tasks, or other cognitive demands (e.g. executive function) or theory of mind deficit. But since they found that when using the eye-tracking method the group difference between typically developing children and children with ASD was still significance after covarying out the verbal, the nonverbal intelligence and even the verbal FBT. However, the results leave two possible explanations of the data; (i) that children with ASD were not motivated to monitor other's mental states, since spontaneous looking behavior was measured or (ii) that children

with ASD do have theory of mind deficit, and even they do pass the verbal FBT it could be because they are using some verbal compensatory strategy, as Happé suggested. However, since this study tested children below the verbal mental age of 7, could not get closer answering this question.

4.2.2. Language in autism

As we already saw verbal ability, or at least some aspects of it, and the performance on FBT are highly associated in autism. Therefore it is crucial to overview the different aspects of language ability in autism.

Language ability in ASD is very heterogeneous. At one extreme of the spectrum we find individuals who do not use language – neither verbal nor nonverbal – therefore they are functionally mute. At the same time, at the other extreme of the spectrum, there are individuals who can acquire formal language at a similar or even at the same level as typically developing individuals. The majority of individuals with ASD show language delay during development and some deficit in different aspects of language ability. Thus, is very important to specify the group of people with autism, when we talk about their language abilities, since they can be very different depending on the severity of the symptoms and language ability. I am going to discuss the ability of people with Asperger syndrome (AS), high functioning people with autism (HFA) and low functioning ones (LFA). Another feature of language ability that is associated with almost all ASD individuals who have at least some language is that their comprehension is worse than their expression, due to echolalia or the repetitive use of grammatical forms (Boucher, 2003).

4.2.2.1. Pragmatics

The deficit in pragmatics is one of the diagnostic criteria of autism, but the severity of its impairment can vary depending on the severity of the symptoms. However, as already discussed above the Dissertation concentrates on the formal aspects of language; mainly semantics, syntax and a little bit of phonology.

4.2.2.2. Lexical-Semantics

Individuals with AS have mild semantic impairment, which manifests in their difficulties in understanding and using non-literal language, including metaphors, irony, jokes or word play, despite their normal vocabulary (e.g. Boucher, 2003). Individuals with HFA have moderate impairment in semantics, in particular with abstract terms e.g. with deictic terms, that change

their reference according to the speaker (e.g. personal pronouns, I or you), place (e.g. here, there) and time (now, tomorrow). Mental state words represent a special group of words in ASD. Just like in typically developing children a strong association was found between mental state words and false belief understanding. Therefore the results that found that mental state words are very rare or completely missing from spontaneous speech of individuals with HFA or especially with LFA are not surprising (Tager-Flusberg, 1992; Ziatas, Durkin, & Pratt, 1998). Moreover, correlation was found between the comprehension of mental state words and the performance on FBT (Ziatas, et al., 1998) In LFA the severe deficit of the above mentioned language impairment can be observed with a lower vocabulary. Several other phenomena can be found in HFA and mainly in LFA individuals' language, that are present in early typical language development too, but in ASD they persist well beyond early development, such as neologisms (made-up or nonsense word) or echolalia (immediate or delayed repetition of words or phrases that are not necessarily appropriate to the current context). These unusual lexical patterns are not specific to autism. As I mentioned most of them can be found in typical development and also in other developmental disorders. However, the frequency and persistence of these atypicalities are distinctive in autism. As it has been probably recognized though, most of these semantic deficits reflect the application of the concepts for communicative behavior rather than a selective semantic deficit (e.g. Tager-Flusberg, 2000). In addition, communicative behavior is impaired in ASD by definition.

Another interesting line of semantics in ASD is the performance of these people performance on the FBT. As I referred to it earlier, Happé (1995) and Fisher (2005) found strong relation between ASD individuals' vocabulary and their performance on verbal FBT. More specifically it was found (Happé, 1995) that children under the verbal mental age of seven (measured by the PPVT, receptive vocabulary) do not pass FBTs. Happé argued that these results might reflect a verbal compensatory strategy that these children use to pass these FBT.

4.2.2.3. Phonology

Phonology is the area in the language ability which appears to be relatively intact in structure (Boucher, 2003; Wilkinson, 1998). On the other hand, however, the prosody of speech has been reported impaired in basically all of its components; intonation, volume and rhythm. But again, the manifestation of atypical prosody can widely differ between individuals with ASD. Among the ASD individuals who acquire at least some language we can find some whose speech is monotone or "wooden", while others' are unusually loud or have a singsong vocal

quality (Fay and Schuler, 1980). Another interesting aspect of prosody can be found in echolalia. Sometime echolalic repetitions retain the exact same prosody of the original sentence or phrase however the context is different from it and therefore "somewhat" inappropriate. For instance, a child repeats her mother question "Do you want some milk?" using question-intonation to request some milk, however it has also been reported that some child use statement-intonation (with falling at the end) for the same request. Therefore, it has been argued (Prizant, 1983) that these echolalic repetitions do have communicational functions, although their prosody (question-intonation to request) or syntax (question with statement-intonation to request) is not correct. Only a very few studies focused on the comprehension of prosody. However in a recent study it was found that adolescents with HFA do not use prosodic cues to disambiguate syntactically ambiguous sentences (Diehl, Bennetto, Watson, Gunlogson, & McDonough, 2008). Thus, there is some evidence that individuals with ASD have problems not only with production but also with comprehension of pragmatics.

4.2.2.4. Syntax

Syntax has been reported intact in people with AS. Although individuals with HFA and LFA show at least some level of impairment in their syntactic skills measured by standardized tests (Jarrold et al., 1997; Tager-Flusberg, 1994), this deficit it not syndrome-specific relative to other individuals with developmental disorders (Boucher, 2003; Wilkinson, 1998). (Tager-Flusberg & Sullivan, 1994) were the first who turned their attention to syntax in addition to lexical-semantics and IQ in relation to theory of mind in autism. They found that the strongest predictor of performance on false belief tests was the measure of syntactic comprehension measured by a standardized measurement (CELF-R). There is one aspect of syntax however, which recently got into the focus of interest regarding the syntactic abilities of ASD individuals, and this is complement syntax.

4.2.2.4.1. Complement syntax

As we saw above Jill de Villiers found that the acquisition of sentential complements are the prerequisites of the emergence of theory of mind, more specifically false belief understanding (e.g. de Villiers, 2007). Two studies focused on the same phenomenon on children with ASD (Tager-Flusberg, 2000; Lind & Bowler, 2009). As we saw language has an even more special, closer relation to FB understanding in individuals with ASD (or at least in those who have at least some language). Thus, if de Villiers' hypothesis is true, we can expect the same or even

stronger relation between sentential complements and FBT. In the Tager-Flusberg study 20 older children and adolescents with autism were tested on a Complements in wh-questions task, on 2 standard location change false belief understanding tasks and on other language tasks; PPVT and Sentence Structure tasks. Unfortunately no further information was given about their IQ or language abilities. These children and adolescents were matched to 20 children and adolescents who suffered from learning disability (the paper refers to them as mentally retarded), therefore we can conclude they were not high functioning individuals with autism – they were matched on age, IQ and language ability. The task was used to measure complement syntax was the Complements in wh-questions (de Villiers and Pyers, 2002). Short stories were presented to the participants and then wh-questions were asked such as:
The little girl went shopping one afternoon but she was very late going home. She went a short way home over a fence but she ripped her dress on the wire. That night when she was in bed she told her mom, "Look I ripped my dress this afternoon!"
When did the girl say what she ripped?
Children around the age of 3-4 years were shown to be able to interpret "long distance" wh-questions such as *When did he say he hurt himself*, since it allows the further interpretation: *He said he hurt himself when* in which the wh-word (when) originates in the lower clause just like in an echo question. However, in case of the "short distance" wh-questions such as *When did he say how he hurt himself* children at the same age make a common error by answering the medial question, in this example how he hurt himself instead of extracting the complements from the complex wh-question. Typically developing children do not tend to answer these short distance sentences under the age of 5 (de Villier, Roeper, & Vainikka, 1990). Therefore the reason why de Villiers did not find any relation between the wh-questions and false belief understanding was probably because it develops later than FB understanding, but it might be an appropriate task to test the mastery of complement syntax in children with autism, who tend to pass FBT at a later age if at all. Tager-Flusberg's results supported de Villiers' hypothesis, she found that it was the performance on the "short distance" sentences, which predicted both groups performance on the FBT, as opposed to the performance on PPVT and Sentence Structure tests, which did not predict this performance. Also, in both groups those who passed the FBT gave significantly more correct "short responses" than those who failed, but this did not hold to the "long distance" wh-questions. In a very recent study, however Lind and Bowler has criticized the Tager-Flusberg study for two reasons: (1) there was no direct comparison of the ASD and control group, whether in the ASD group there is stronger association between complements and FB understanding and (2)

because of the inappropriate statistical analyses, because they used regression analyses on a relatively small sample size (n=20). Lind and Bowler used a bigger sample size indeed (n=48), however this made their sample extremely heterogeneous. In their ASD sample there were both children and adults, low and high functioning individuals and individuals with AS. In their control group they mixed typically developing individuals and individuals with leaning disabilities (around 50-50%), again both children and adults. Clearly, when a developmental relationship is in question, the age of the participants is crucial, not to mention the relative homogeneity of the studied atypical populations. (More about the problem of matching comes later). They found strong and significant correlation in the ASD group between complement syntax and FBT even when verbal mental age was partialled out. The relation in the typical and learning disabled mixed group was still significant but weak. Also note that the wh-questions applied in the Tager-Flusberg paper might be different across languages, for instance the Hungarian tally of the mentioned long distance wh-question is *Mit mondott, mikor sérült meg* (translation word-by word would be What did he say when he hurt himself*) which also contains 2 wh-words, mit (what) and mikor (when) interrogatives, which might cause different outcome in a study.

4.2.2.5. Is language intact in ASD?

As we saw, individuals with ASD have most of their linguistic problems are somehow related to the social context (e.g. deictic terms, echolalia etc). Thus, it is also a plausible question whether standardized paper-pencil language tests (e.g. Peabody, CELF, TROG etc) are appropriate methods to measure the language ability of this population. However, a recent study found strong correlation between these standardized tests and measures of spontaneous speech (e.g. MLU, or Index of Productive Syntax) suggesting that the standardized tests reflect validly and reliably the language ability of this population (Condouris, Meyer, & Tager-Flusberg, 2003).

4.3. Specific language impairments

Language impairment can be associated with many developmental disorders, individuals with SLI selectively or at least primarily have language deficit – depending on the cognitive developmental model we apply. The criterion of an SLI diagnosis is that the language

impairment is present in the absence of any cognitive, sensory, neurological or social-emotional deficit. The lack of cognitive deficit usually interpreted as the non-verbal IQ falls into the normal range as opposed to the verbal IQ which is considerably lower, meaning it is usually 2 SD lower that the average. The prevalence of the disorder based on American assessments is approximately 7% (Leonard, 2000) and the female-male ratio is 1:3 affecting more males than females (Temple, 1997).

Children with SLI are not a homogeneous group. The deficit of language can be manifested in numerous ways in SLI; different aspects of language can be affected in language (or language faculty): phonology, semantics, grammatics or pragmatics. These aspects can be impaired in different degree, moreover either the comprehension (receptive deficit) or the production (expressive deficit) or even both can be impaired in certain cases. However, what really makes the deficit heterogeneous is that these aspects affected in different degrees can create the most divers linguistic pattern or profile (e.g semantics is relatively good but grammatics is bad). This heterogeneity can be even more enhanced, since the linguistic profile can further change during development, which can even lead to the change of the primary linguistic symptom (e.g. the certain language aspect) (Conti-Ramsden, Crutchley, & Botting, 1997; Csépe, 2005; Hahn, Győri, Várnai, & Sajó, 2006). There are two important implications of heterogeneity; (1) the ambition to create more homogeneous subgroups within the impairment and (2) it is very difficult, or even impossible to find a single causal model, which is able to explain the complete variability of the deficit. Two methods are available to create subgroups within SLI; with statistical methods and with clinical methods, both of which are based on testing these children on various neuro-cognitive tests. Unfortunately until now, there is no consensus what these subgroups are, moreover there is no consensus regarding the number of the subgroups either (Bishop, 1997; Van der Lely, 2005; Rapin & Allen, 1987). But even if there were subgroups based on consensus an additional problem would still remain, namely that belonging to a subgroup is not a stable feature of the person's language impairment, sometime the person switch among subgroups (Conti-Ramsden, Crutchley, & Botting, 1997).

As mentioned above, the other implication of the heterogeneity at least in part is that there is no consensual (cognitive) explanatory model of SLI. Without reviewing these models[9] I will focus on the question whether language is selectively impaired in SLI or other cognitive abilities, in this concrete case theory of mind is impaired too. The question is plausible since

[9] Since it is not the aim of the Dissertation to take a stand on the specificities of the linguistic deficit within SLI (see e.g. Pléh, 2007; Leonard, 1998).

if one of this ability is necessary to acquire the other, as some of the above discussed theories suggest, language ability could not be selectively impaired.

There are relatively few studies in the literature that concentrate on this specific question. The early studies tested ToM ability in children with SLI for instance with more emphases on the autism group and used the SLI group as a control group and concluded that children with SLI pass FBT in spite of the language deficit while the majority of ASD children did not (Leslie & Frith, 1988; Perner, Frith, Leslie, & Leekam, 1989). These results suggested that language deficit itself is not responsible for ToM deficit in ASD children.

Only recently got the ToM ability of SLI children into focus of a few studies concentrating on a more exhaustive studying of the developmental process of ToM in SLI. As a result of these more exhaustive studies the findings of this question became somewhat controversial. Some studies found that the acquisition of ToM has been delayed in children with SLI (Farrant, Fletcher, & Maybery, 2006; Miller, 2001; Gillott et al, 2004), by approximately 12-18 months (Tucker, 2004). In these studies ToM ability was accessed by a variety of tasks and not only by FBT. The controversy between the more recent and the older studies can be resolved since the older studies' SLI groups were significantly older (around the chronological age of 8-9). The later results that found delay in ToM development raise the question whether these data reflect a real impairment in underlying ToM ability or rather result from difficulties in managing the linguistic complexity of these tasks. According to my knowledge there has been only one study that systematically studied the relation between the performance on FBT and the linguistic demand of FBT on children with SLI. Miller (2001) presented the usual verbal false belief story to the children and then varied the linguistic demand of FBT by using different test questions; (1) where does the puppet *think* the toy is (2) where will the puppet *look* for the toy and (3) *Show* me what the puppet will do (or what happens) by giving the puppet to the child, and arguing that they are decreasing in difficulty in this order. They found that the 'think' condition is more difficult for children with SLI than either (Astington et al, 2005) the 'look' or the 'show' conditions. Also, children with SLI performed above chance in these conditions but not he 'think' condition, in which they performed similar to the verbally matched (typical) control group, while the age matched control group performed above chance in all of the conditions. However, there are several limitations of the study, first, the very small number of participants, only 9 children with SLI took part. Second, the results mentioned above are based on the cumulative data of three sessions where the same tasks were administered. After analyzing the data of only the first session, no such differences were found. Thus, it might be that children with SLI improved on these tasks due to practice –

however, it would not explain why they did not improve in the 'think' condition. And finally the results are probably due to the very high proportion of invalid trials in the show condition, most of which was due to "either experimental error, or a failure to respond in a relevant manner" (p. 81). In fact, more children with SLI passed the 'think' condition then the 'show' condition, but while most of the rest children failed the 'think' condition, the same amount of children produced invalid trials in the 'show' condition. It is also important to emphasize that these FBTs were presented verbally, which still represents a considerable linguistic demand for children with SLI.

As mentioned earlier in the autism part, there is one study that tested FB understanding of children with ASD and children with SLI on a completely nonverbal test. I also mentioned, that due to matching, the SLI group's chronological (and nonverbal mental age) were too high (around the age of 8) to test the question of delay in ToM acquisition, in consensus with other studies who applied verbal FBT at this age, they found that children with SLI performed on the nonverbal FBT well above chance, around 70%. The question, however, how younger children with SLI would perform on a nonverbal FBT remained open.

4.4. The problem of matching

Selecting the appropriate control group to a certain atypically developing group is crucial in terms of the results and especially in terms of interpreting these results. As we saw above, this can be very challenging and can be the source of major criticism. There are several questions that have to be considered when selecting control group(s), however there is no consensus in the literature in terms of what a good control group(s) is. Usually the first question to consider is whether the control group should be typically or atypically developing. In general, using an atypical control group in addition to the typical control group is crucial to find out whether the deficit is a deviance – atypical pattern of development – or a "just" a delay – typical pattern of development. Selecting the appropriate atypical control group can be challenging too. For instance, it has argued recently that Down syndrome (DS) is not an ideal atypical control group – although many older studies compared their atypical group to individuals with DS – due to its uneven cognitive profile e.g. specific strength in attentional, social and emotional abilities that may be associated with higher abilities in ToM tasks. Hence, recently a less specific group with mental retardation (or in other words with learning disabilities) with unknown etiology has been used. If the atypical control group has been identified, we still need to decide whether we should match groups or individuals, the matching should based on chronological or mental age and/or based on other ability (e.g. language) and also what

instrument to use to measure these abilities. For instance it has been found that results can be affected by the instrument used to measure IQ (PPVT, Raven, Wechsler) (Mottron, 2004). Another problem of matching based on full-IQ or mental age is that IQ tests are very complex and a lot of developmental disorders, including ASD have their strengths and weaknesses in different subtests. Therefore two groups having the same IQ points can have very different underlying abilities (Jarrold & Brock, 2004). Focusing on ASD, separating verbal IQ from nonverbal IQ does not solve the problem either, since individuals with ASD have higher nonverbal IQ than verbal IQ (Joseph et al, 2002), and even within the verbal and nonverbal IQ further unevenness can be found. An alternative approach of the problem is to create a so-called matching design in which a task is explicitly designed to share most of the features of the task assesses the ability in question (independent variable) except for the critical feature (Jarrold & Brock, 2004). An example of such a matching design could be the application of the true belief task when false belief understanding is tested, since the only difference between the two tasks is the attribution of false belief. However, as we saw above this assumption has been criticized (Roth & Leslie, 1988; Riggs & Simpson, 2005) since in TBT the character's (true) belief is identical to the state of reality as opposed to the FBT where they are different which adds at least an extra inhibition, and not only false belief attribution to the FBT. As we saw in the TBT part it, until now there is no TBT in which the only difference between that and FBT is false belief attribution.

Chapter 5. Studies

5.1. Study 1: A New Nonverbal False Belief Test: the performance of typically developing children and children with developmental language impairment

5.1.1. Background

The development of a new nonverbal FBT had several motivations.
(1) First is to create a FBT that is able to test children with low or no language ability in the critical age, so around the age of 4.
(2) we saw that decreasing the linguistic demands of the FBT can result in better performance in earlier age, so it is possible that the verbality of the FBT masks the already existing FB understanding ability under the age of four.
(3)a third motivation is a theoretical one: without a nonverbal FBT it is very problematic to get a finer picture about the developmental relations of language and ToM, or more specifically false belief understanding.
(4) as we saw it in Chapter 2 there are already a few nonverbal FBTs, however we also saw that all of them have their shortcomings. The aim of this new test is to overcome at least some of problems, disadvantages of the already existing nonverbal FBTs. Thus the aim was to create a FBT that is completely nonverbal as opposed to Astington or de Villiers test, does not require former training and therefore it is easy to administer even with children with disabilities, as opposed to the test created by Call & Tomasello (1999), it requires a complex goal-directed intentional behavior as a response instead of gaze response (Clements & Perner, 1994; Onishi & Baillargeon, 2005, Southgate et al, 2007), it is sensitive in the critical time of false belief understanding in typical development, around the age of 3-6, and most importantly it is a valid test of false belief understanding.
The basic idea of the test is that children are presented with short and simple stories acted out with puppets, and the children have to finish these stories. The first few stories are tuning situations during which the children understand that they are presented stories with pretend play features and at a certain point they will get the control over the main character (puppet) and finish the story. The other aim of the tuning phase is to gradually increase the complexity of the tasks, to make sure that the children can understand these more and more complex situations by giving an adequate behavioral response. By the time of they finish the tuning

phase they can adequately finish true belief situations. Once children passed the tuning phase they are presented with a demonstration. During the demonstration children are presented with a false belief situation, which is acted out by the experimenter completely. After the demonstration, two false belief test situations and one true belief test situation is presented to the children. They are expected to finish these stories by leading the first puppet to the correct hiding place (which is in case of the FBTs are the empty places). It is argued that during the test phase the only additional skill is needed compared to the tuning phase is false belief understating. (Detailed description of these situations can be found in 5.1.5.)

5.1.2. Hypotheses and questions for typically developing children

Study 1 was designed to test the following hypotheses and questions:

(1) The new nonverbal FBT is a valid test of false belief understanding
(2) it is hypothesized that – since the new nonverbal FBT does not require more abilities or a higher level of these abilities – children can pass the test at least at the same age as the verbal FBT. Therefore we expect children to pass the test around the age of 4.
(3) whether similarly to the verbal FBT there is developmental shift in the performance of the nonverbal FBT; 4-year-olds performance is significantly better than 3-year-olds' performance.
(4) whether the nonverbal FBT has a relation to the language tests (Peabody & TROG-H), while such a relation is expected between the verbal FBT and the language tests.

5.1.3. Hypotheses and questions for children with Developmental language impairments

(1) it is expected that children with DLI show a delay in the performance of the verbal FBT
(2) whether children with DLI has a delay in the performance of the nonverbal FBT
(3) whether a relation will be found between the two FBTs (verbal and nonverbal) and the verbal tests

5.1.4. Participants

The participants were 48 typically developing children and 22 children with developmental language impairment from Budapest, Hungary. The two groups of children were not matched to avoid ceiling or bottom effect, but they were selected from the relevant age ranges based on the literature (e.g. Baron-Cohen, 1985; Wellman, 2001; Miller, 2001). As typically developing children and children with developmental language impairment pass false belief tests at a different ages (around the age of 4 and around the age of 5;6-6, respectively), age matching would have caused either a ceiling (ages of typically developing children matching to the ages of children with DLI) or a bottom effect (ages of children with DLI matching to the ages of typically developing children). Due to the lack of matching the two groups were analyzed respectively and no between-subjects comparisons were made. Parents signed consent forms and children received a small gift for participating.

Typically developing children

In the typically developing group the participants were 16 3-year-old children, 5 boys and 11 girls (M = 42.83 months or 3;6 ages, range = 3;0 – 3;11), 16 4-year-old children, 8 boys and 8 girls (M = 55.0625 months or 4;7 ages, range = 4;0-4;11) and 16 5-and 6-year-old children, 6 boys and 10 girls (M = 68.19 months or 5;8, range = 5;0-6;6). The children came from diverse social-economic backgrounds and were all native Hungarian speakers. They were recruited from two kindergartens (Győri Apor Vilmos Katolikus Iskolaközpont Óvodája; Budapest, VIII. ker. Vajda Péter utcai Óvoda). Two children were excluded from the sample because of having some developmental disorders. Thus, the total of 50 children were seen.

Children with Developmental language impairments

In the developmental language impairment group there were 14 boys and 7 girls, ranging in age from 5;0 years to 7;7 years (M=6;5 years, SD=8.4 months). The range of the Leiter (nonverbal) IQ was 68-109 (M=90, SD=11.2) and the range of nonverbal mental age was from 4;3 years to 8;7 years (M=5;6 years, SD=13 months). Children were enlisted from the Speech Therapy Kindergarten of ELTE Bárczi Gusztáv College for the Handicapped; among the data of the 34 investigated children we excluded those who had been diagnosed with

learning disorder, mental handicap (IQ<70)[10] or attention disorder beside the language disorder and whose language disorder was articulation kind of. We have to emphasis that due to the lack of standardized Hungarian psychological tests and especially language tests, our sample was still very heterogeneous considering to the nature and seriousness of the developmental language impairment and the level of other abilities (list of diagnosis in Hungarian see in Appendix 1).

5.1.5. Materials

Two false belief tests, a verbal and a nonverbal one, and two language tests were given to each child.

The new *nonverbal false belief understanding task.*
The logic of the nonverbal FBT was that during the tuning phase the child was tuned to finish the story that the experimenter started by acting out with puppets. Therefore there is an obvious pretend play component of the task. Before the testing phase a demo was presented to the child, in which a false belief understanding situation was acted out by the experimenter to avoid the child's assumption that the goal of the task is to adequately finish the story, which might mean to find the hidden object in the FBT. In the testing phase then, false and true belief situations were presented and the child was encouraged to continue the story from the point when the first puppet comes back. The application of the TBT was necessary to ensure that the child does not learn a certain response schema during the demo (e.g. lead the puppet to the empty place). The child's behavior, where she/he led the puppet was coded.

1. „Cube carrying": there are 4 cubes on the table next to each other, about 20 cm far from them a man figure is standing, holding a wheelbarrow in his hands. The puppet starts to go to the cubes and puts one into the wheelbarrow, then goes back to his starting point. There he lays the cube down and goes back for another one. After carrying the second cube over, the experimenter pushes the man with wheelbarrow in front of the child and encourages the child to continue.

[10] Since the international literature's criteria of SLI is a minimum IQ of 85, the calculations were also made with this smaller group of children, however probably this affect only 5 children, the results were the same as with the larger group presented in the dissertation.

1. B "Watering flowers": if the child could not pass the *1. " Cube carrying"* task (it has never happened yet) the experimenter puts 6 little artificial pot flowers next to each other on the table. A puppet comes in with a watering can in her hand. She goes to the first flower and waters it, then to second and to third and waters them too. The experimenter stands the puppet in front of the child and encourages the child to continue. In case the child doesn't continue the task (doesn't continue watering the flowers) we don't skip to the next task, the child is excluded from the test.

2. "Swinging": the experimenter puts a swing in front of the child in which a puppet is sitting. Another puppet starts to push the swing from the back. After about 3-4 pushes the puppets change places so the puppet, who was sitting in the swing is the pusher and the other puppet sits into the swing. After a few more pushes the experimenter gives the pusher puppet to the child and encourages the child to continue.

3."Tea party": the experimenter places a little toy table and 2 toy chairs in front of the child and on one of the chairs a puppet is sitting. Next to the toy table there are plates, cups a teapot and a spread. The experimenter starts to set the table with the other puppet and after placing the spread and 1-2 cups on the table the experimenter gives the puppet to the child and encourages the child to continue.

4. "True belief: Sally-Anne": the experimenter puts a basket and a box in front of the child behind of both a puppet is standing. The puppet standing behind the basket has a ball and puts it into the basket then leaves. In the meantime the other puppet takes the ball out of the basket and plays with it for a short time then puts it back into the basket and leaves. Then the first puppet comes back and at this point the experimenter gives the puppet to the child and encourages the child to continue. This was the last task of the tuning phase and if it was passed we had good reason to suppose that the child acquired an action schema: that the experimenter starts an action and he/she has to continue it in an adequate way. The next phase was the test phase.

5. "Demo: False belief, garages": before presenting the demo, the experimenter signed to the child that now the experimenter was going to play the whole story. There were two toy garages in front of the child for about 20 cm far from each other and 1-1 puppet was sitting next to each of them. One of the puppets had a car and parks it in his own garage then leaves.

Then the other puppet comes out of the garage with the car and parks it in her own garage and leaves. The first puppet comes back goes to his own garage and gets surprised (the experimenter signs it with facial expressions and with intonations like "Oh!") when he didn't find his car in there. He starts looking for the car at more places then finally he finds it in the other garage. This task was enrolled to avoid the child's possible misinterpretation of the former tasks; that the goal of the game is to find the object. With the demon it was demonstrated that this is also a good and possible continuation. If the child was not able to attribute false beliefs yet then it was very unlikely to transfer this knowledge to other tasks.

6. "Test 1: False belief: Sally-Anne": the standard Sally-Anne story was presented for the child without any verbal comments. The story is identical to task 4. "True belief: Sally-Anne" except that after the first puppet leaves the other puppet takes the ball out of the basket and puts it into the box (replacement) and leaves. The first puppet comes back and it is given to the child to continue the story. If the child led the puppet to the basket the child passed the task if she/he led it to somewhere else e.g. to the box she/he failed.

7. Test 2: True belief; cooking": the experimenter places two pans on a table, a blue and a white one in front of the child. One of two puppets puts a piece of chocolate into the blue pan and starts to stir it then leaves. At the meantime the other puppet who was watching from the corner goes to the blue pan takes the chocolate out, tastes it and puts it back to the blue pan then leaves. The first puppet returns, and the experimenter gives the puppet to the child and encourages the child to continue. If the child leads the puppet to the blue pan then she/he passes the task, if the child leads it to the white pan she/he fails. Adding the true belief task to the test helps to avoid a possible strategy created by the child e.g. she/he has to lead the puppet to the place where is nothing. With a strategy like this the tasks with replacement could be passed without using false beliefs.

8. "Test 3: False belief; going to sleep": the experimenter places a bed and two chests, a small and a big one in front the child. One of the puppets has a teddy bear and puts it into the big chest then goes to sleep to the bed. While this puppet is sleeping the other puppet, who was sitting in the corner takes the teddy bear out of the big chest and puts it into the small one then leaves. The other puppet wakes up and the experimenter gives the puppet to the child and encourages the child to continue. If the child leads the puppet to the big chest she/he passes the task if to the small chest or somewhere else she/he fails it.

In this test, as we mentioned two location change false belief tasks and one true belief task were presented. Both raw scores (0-3) and pass/fail (1-0) scores were used during analysis. Children were credited with passing the nonverbal false belief test if they passed all the three belief tasks. The motivation of this strict criterion of passing the nonverbal task was that since there was no control question in the nonverbal test we wanted to make sure that the child not just simply applies a certain type of behavioral schema to finish the story (e.g. Lead the puppet to the empty folder), but attributes beliefs (true or false) to the puppets.

Verbal false belief understanding tasks. Two location-change and one unexpected content task was used. One of the two location-change task was the standard Sally-Anne task, the other was a chocolate cooking story. The schemas of the two stories were the same. The first character puts an object (ball or chocolate) into a covered place (basket or cooking pot) while the second character is watching her. In the absence of the first character, the second character replaces the object (ball or chocolate) into another covered place (box or different colored cooking pot) and leaves. When the first character comes back the child is asked where she is going to look for her object (ball or chocolate) first. The child had to answer both the test question and the 3 control questions in order to pass the task. The control questions were a reality question (were the object really is), a memory question (where the first character put the object in the beginning) and an identity question, which checked whether the child knew which character was which. The unexpected content task was the Smarties task in which the Smarties tube contained a pencil. Children were asked both about another person's false belief (what would another person think is in the box) and about their own false belief prior to the tube being opened. They were only credited with passing the false belief questions if they also passed a reality control question (what is in the box now). If children could not pass the false belief question but they could answer all of the control questions; the memory, the reality and the identity questions correctly they failed the task. If they could not answer any of the control questions correctly then their performance was not evaluable and children were not given any scores. Both raw scores (0-3) and pass/fail (1-0) scores were used during analysis. In case of the pass/fail scores the child had to pass at least two of the three verbal false belief tests to credit him/her with passing the verbal false belief test. By using this criterion we could avoid that the performance at chance level would be enough. This level of criterion was motivated both by the criterion of other studies (see Wellman et al, 2001 for review)) and by the fact that one of the three false belief tasks was the Smarties (unexpected contents) task which has been

argued to be more difficult for the children (Astington, 1994) than the location-change false belief tasks.

Thus, stricter criteria were used in the nonverbal false belief test, than in the verbal one. As we argued, the reason of this is the lack of control questions in the nonverbal FBT. Nevertheless, we couldn't completely eliminate the differences of the chances to pass the test. In the nonverbal test, the chance to lead the puppet to the right place is 50% in each trial, so the chance to pass all the three trials of the test, and therefore the test itself, is 12.5%. However the chance to pass one trial of the verbal FBT is 6.25% due to the 3 control questions (50% to answer correctly each) so the chance to pass the complete verbal FBT is 0.39%. This is definitely the disadvantage of the nonverbal FBTs in general. On the other hand, however it is also more difficult to fail the verbal FBT than the nonverbal FBT, since if children fail the false belief question and they also fail any of the control questions they performance is not evaluable.

Test for Reception of Grammar (TROG). The Hungarian version of TROG (Bishop, 1983) (TROG-H) was used as a test of development of receptive grammar. This task involves presenting the child with four pictures, and asking them to indicate which picture goes with a sentence containing a grammatical construct. Items are arranged in blocks of 4, all of which test the same grammatical construct, and a child is considered to have failed a block if they fail a single item within the block. If the child fails 5 consecutive blocks the test is discontinued. The number of correct blocks was used as the index of the development of grammar, because the standardization of Hungarian version is still in progress[11].

Peabody Picture Vocabulary Test (PPVT). Receptive vocabulary was assessed using the Hungarian version of Peabody Picture Vocabulary Test (Csányi, 1974). In this test, children are shown four pictures and asked to point to the picture that best tells the meaning of a word. Items become progressively harder. The test is discontinued when a child makes 6 errors in a group of 8. In the lack of a complete and updated Hungarian standard of the PPVT, the raw scores were used as the index of vocabulary.

[11] The development of the Hungarian version is in progress by Ágnes Lukács and Miklós Győri (BME KT Department, Budapest)

5.1.6. Procedure

All children were tested in two sessions, individually in a quite room in their kindergarten. Each session took approximately 30 minutes. On both sessions one language and one false belief task were alternated, and the order of each language and false belief tasks were counterbalanced. The two sessions were finished in one month in order to ensure the child is at the same developmental level during the two sessions.

5.1.7. Results

Preliminary results showed no order or gender effects; therefore, these variables were collapsed in subsequent analyses.

5.1.7.1. Results on typically developing children

Validating the nonverbal FBT

To strengthen the results of the new nonverbal FBT we used both raw scores and pass/fail scores of the nonverbal and verbal FBT in validation.

Figure 8 presents the percentage of children who passed the verbal and nonverbal false belief tests as a function of age[12]. 3-year-old children did not pass at above chance levels in either test. Only six children (37.5%) passed the nonverbal FBT, and only four children (25%) passed the verbal one (Binominal, $p>0.05$). In contrast, 4-year-old children did pass both verbal and nonverbal FBT above chance level; eleven children (69%) passed the nonverbal FBT and 11 children (69%) passed the verbal test, however not significantly above chance (both Binominal, $p>0.05$). And finally, 5-year-old children passed both verbal and nonverbal tests significantly above chance level; fourteen children (87.5%) passed the nonverbal (Binominal, $p\leq0.01$) and 16 children (100%) passed the verbal FBT (Binominal, $p\leq0.001$). Thus, a developmental trajectory of false belief understanding was found between the ages of 3 to 5, and these trajectories of the verbal and nonverbal FB understanding described very

[12] Since I used different pass/fail criteria for the verbal and nonverbal FBT, here we used the pass/fail (dichotom) scores.

similar pattern/span.

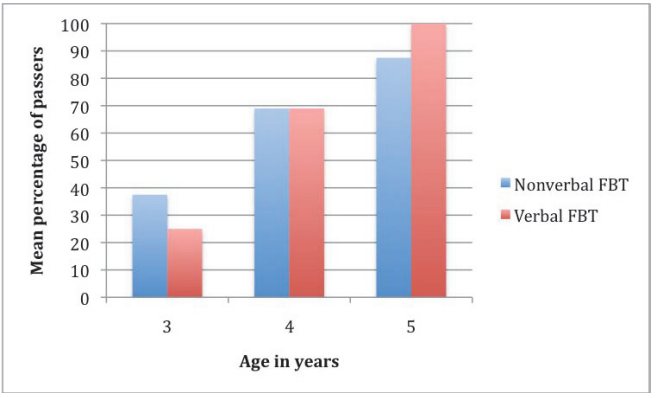


Figure 8. Percentage of children passing the nonverbal and verbal false belief tests as a function of age.

These results are important in the validating process of the nonverbal FBT since they suggest not only that children pass the nonverbal FBT at the same age as they pass other verbal and nonverbal false belief tests (see e.g. Call & Tomasello, 1999; Milligan et al, 2007), but also because we found similar pattern in the development of the performance of verbal and nonverbal FBT, suggesting similar mechanisms in the background.

A 3x2 (age x FBT) ANOVA with age (3-,4-, and 5 years) as a between subject factor and FBT as a within-subject factor. The analyses revealed a significant main effect for FBT $F(1, 45) = 13.47$, $p \leq 0.001$. Regardless of age, children's performance on the nonverbal FBT is better compared with the verbal FBT. In addition there was a significant main effect for the between-subject factor age ($F(1, 45) = 16.93$, $p \leq 0.001$, indicating that overall, 5-year-olds performance on the FBTs was better than younger children's performance. The effect of interaction was also significant $F(1, 45) = 3.16$, $p \leq 0.05$). Subsequent analyses comparing the mean score of the nonverbal FBT and the verbal FBT within each individual age group revealed that the performance on the two tests was significantly different at the age of 3 years ($t(15) = 4.04$, $p \leq 0.001$) but not at 4 -or 5 years (4-year-olds: $t(15) = 2.07$, $p > 0.05$; 5-year-olds: $t(15) = 1.86$, $p > 0.05$), suggesting that 3-year-olds performance on the Nonverbal FBT is significantly better than on the Verbal FBT. Further subsequent analyses found that the 4-year-olds performance on both verbal and the nonverbal FBT is significantly better than the 3-year-olds', performance (verbal FBT: $F(1, 30) = 11.81$, $p \leq 0.05$; nonverbal FBT: $F(1, 30) =$

5.75, p≤0.05) but there is no such difference between the 4-and 5-year-olds (verbal FBT: F(1, 30) = 2.3, p>0.05; nonverbal FBT: F(1, 30) = 0.83, p>0.05)[13]. These results again strongly resonate with the literature; there is a developmental shift around the age of four in theory of mind development (e.g. Wellman et al, 2001). Similar analyses were made with the pass/fail scores and similar results were found. Significant increase was found in the verbal FBT between the ages of 3 and 4 (χ^2 (2, N = 32) = 6.15, p ≤ 0.02), however this difference was only a tendency in the nonverbal test (χ^2 (2, N = 32) = 3.14, p ≤ 0.08). No significant differences were found between the ages of 4 and 5.

An important step of validation is to investigate the relation of children's performance in the two measures (verbal and nonverbal) of false belief understanding. Raw scores (0-3) were used to correlate the values of the two tests. Pearson correlation revealed a significant correlation between the two tests (r =0.44 , p ≤0.01 , df = 46). However, when partial correlation was used, controlling for age, the correlation was not significant and weak (r = 0.2, p > 0.05). In spite of the similar trajectory in development why did not the two measures of false belief understanding correlate? There are several possible answers. Maybe the ability we measure with the nonverbal FBT is not false belief understanding. However, we find this explanation very unlikely. We used the same location change paradigm that was used in the verbal version. However, the nonverbal FBT clearly has a pretend component. As we saw, pretend play emerges around the age of 18-24 months that is considerably earlier than false belief understanding, so it cannot be an obstacle to pass the test, and the data also suggest that nonverbal FBT is easier for the 3-year-olds. However, this pretend aspect can be at least partially responsible of the lack of correlation between the two FBTs. Another possibility is that the lack of correlation is due to the different probability of passing the two tasks. This would suggest that if we do not take the control questions into account in the verbal FBT then the two measures would correlate. I checked this alternative, after re-coding the verbal FBT accordingly, still no correlation was found (p> 0.05) between the verbal and nonverbal FBT, suggesting that the verbal control questions did not make a difference of passing the verbal FBT and therefore are not responsible of the lack of correlation between the verbal and nonverbal FBT. However, we cannot exclude another possibility either, that no correlation

[13] Considering that the variables of the FBTs are not typical continuous variables (0-3) a more robust nonparametric analysis was also conducted, and the same significant results were found. The Kruskal-Wallis test revealed significant difference among the three age groups in both the verbal, χ^2 (3, N = 48) = 18.58, p < 0.001 and the nonverbal FBT (χ^2 (3, N = 48) = 9.859, p < 0.008). Further comparisons between the groups revealed significant difference both in the verbal and nonverbal FBT between 3 and 4-year-olds (verbal, Mann-Whitney U = 53.5, p < 0.004, nonverbal, U = 77, p < 0.004). No significant difference was found in any of the two tests between the ages of 4 and 5.

was found because the trials of the two tests were different; the nonverbal test contains one true belief and two false belief tasks (two location change tasks), while the verbal test contains three false belief tasks (two location change and one false identity). Another possibility is that the verbal and nonverbal FBT do measure the same ability (understanding false beliefs) but access it through two different ways and that is why these two tests do not correlate. There are some other findings in the literature where the verbal and nonverbal versions of the same task dissociated (e.g. audiovisual object processing: Hocking & Price, 2009; Kobayashi et al., 2007). Finally, we should consider the 'noisiness' in performance of these tests. It has been also argued (Mayes et al., 1996 – they found poor test-retest reliability at this age, and it was not simply due to the development or experience since a considerable subgroup showed decrease in the performance) in the theory of mind or more specifically in the false belief understanding literature that passing false belief test during development and especially between the ages of 3 to 5 is very fragile and not necessarily consistent. A child who passes one test in one moment not necessarily passes it in the next moment (test-retest reliability). The reason of inconsistency can be different; lack of attention, tiredness, getting bored, and since the ability of understanding false belief is not very stable yet, the child is even more sensitive to these other factors. Therefore, we chose those kids, who consistently failed or passed the tasks in the two false belief tests. We found 27 children (four 3-year-olds, 10 4-year-olds, and 13 5-year-olds) who showed this consistency in the two tasks. The pass/fail scores indicated a significant connection between the two tests (Phi and Cramer V's = 0.516, $p \leq 0.01$) However, it has to be admitted that due to this selection our sample of 27 children did not involve equal number of each age group. Almost 50% of the children were 5-year-olds and this could cause modifications in the results. Nevertheless, this finding strengthens the validity of the nonverbal FBT and suggests that the lack of correlation on the entire sample is, at least partially, due to the children's inconsistent performance on the two false belief tests.

Analyzing the nonverbal test by trials

I also compared the mean percentage of performance on the three trials of the nonverbal FBT both within the entire sample and within the three age groups, but there was no significant different, suggesting that none of the trials was more difficult than the others (McNemar, $p>0.05$ in all cases)

Analyzing the trials of verbal FBT

However comparing the three trials of the verbal FBT revealed that mean percentage of the passers of the Smarties false belief trial is significantly lower than the mean percentage of the Sally-Anne (McNemar $p \leq 0.03$) or the Cooking false belief trial (McNemar $p \leq 0.01$). Subsequent analysis found that for children at 3 years of age the Smarties test was significantly more difficult than the Sally-Anne (McNemar, $p \leq 0.05$) or the Cooking false belief trial (McNemar, $p \leq 0.05$). For 4-year-old children the Smarties trial was more difficult only than the Cooking trial (McNemar, $p \leq 0.05$) and for the 5-year-old children there was no difference among the mean of percentage of passers within the three trials (see Figure 9). The results underpin our reasoning of criteria for passing the verbal FBT.

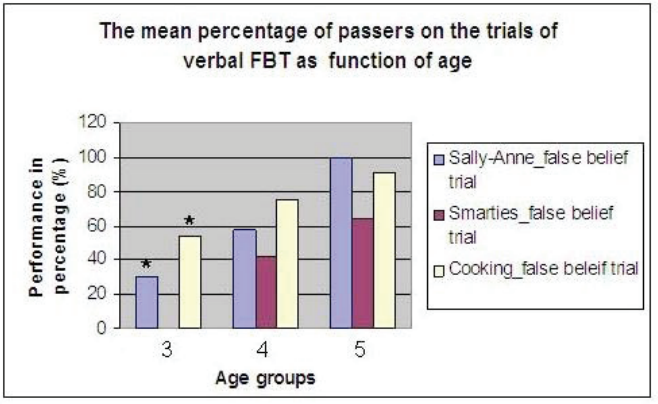


Figure 9. Mean percentage of passers on the trials of the verbal FBT as a function of age.

Relations between false belief tests (nonverbal and verbal) and language tests

In order to investigate the relation between language ability and false belief understanding partial correlations were used, controlling for age, among the false belief understanding tests (using raw scores 0-3) and the language tests measuring different aspects of language ability; vocabulary (PPVT) and grammar (TROG-H). The descriptive statistics of these tests is presented in Table 2.

	3-year-olds	4-year-olds	5-year-olds
Nonverbal FBT (raw scores)	1.88 (1.01)	2.63 (0.62)	2.81 (0.54)
Verbal FBT (raw scores)	0.94 (0.77)	2 (0.97)	2.44 (0.63)
TROG-H	4.44 (2.22)	8.38 (3.95)	12.75 (3.19)
Peabody	44.25 (12.47)	58.94 (15.21)	66.56 (11.28)

Table 2. Descriptive statistics of test performances as a function of age, SD in brackets.

It was found that both PPVT and TROG-H significantly correlated with the verbal FBT (PPVT, $r = 0.51$, $p \leq 0.001$; TROG, $r = 0.34$, $p \leq 0.05$) but none of them with the nonverbal FBT[14] (see Table 3.). Several interpretations of the results are possible (see discussion), but the modality of presentation of false belief tests and the modality of required response - whether they are verbal or nonverbal - seem to be critical in false belief understanding at this age.

	Verbal FBT	PPVT	TROG-H
Nonverbal FBT	0.21 (ns)	0.19 (ns)	0.04 (ns)
Verbal FBT		**0.508 ($p \leq 0.001$)**	**0.34 ($p \leq 0.05$)**
PPVT			**0.4 ($p \leq 0.01$)**

Table 3. Partial correlations of language measures and false belief understanding measures (controlled for age).

5.1.7.2. Results on children with developmental language impairment

Relations between the nonverbal and the verbal FBT in children with DLI

Using pass/fail scores for the verbal and nonverbal false belief tests, 13 children passed the verbal FBT, which is 62% of the children and only 9 children (42%) passed the nonverbal

[14] And the two language measures, PPVT and TROG also correlated with each other ($r = 0.4$, $p < 0.01$)

test. In other words, children with DLI passed the verbal FBT above chance level but not the nonverbal FBT, however none of the tests were passed significantly above chance (Binominal, $p \leq 0.05$). Moreover, the difference between the performance on the verbal and nonverbal FBT is not significant either (raw scores: $t(1, 20) = -1.451$; , $p > 0.05$, with pass/fail scores McNemar test, $p > 0.05$) (see on Figure 12).

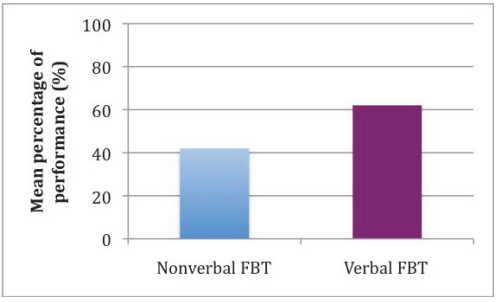


Figure 10. DLI children's mean percentage of performance on the nonverbal and the verbal FBT.

Due to the surprising results I further tested children's performance on the three trials of the tests. In the nonverbal FBT I did not find significant difference in the performance of the three trials. Thirteen children passed the Sally-Anne trial (FB trial), 18 children the Cooking trial (TB trial), and 17 children the Sleeping trial (FB trial). In contrast, in the verbal FBT the Smarties (6 children passed) trial was significantly more difficult than the Sally-Anne (16 children passed) (McNemar, $p \leq 0.01$) or the Cooking trial (15 children passed) (McNemar, $p \leq 0.05$), just like at the typically developing children.

Relations between false belief tests (nonverbal and verbal) and language tests in children with DLI

Age (years)	Leiter IQ (nonverbal)	Peabody	TROG-H
6;5 (8.4)	90 (11.2)	65.28 (17.81)	8.27 (4.22)

Table 4. DLI children's descriptive statistics on the test SD in brackets.

The results on the language tests suggest that in spite of the children's diagnosis, the language ability of these children is considerably good. More precisely, their performance on the Peabody was roughly equal to the 5-year-olds performance and their performance on the TROG-H was roughly equal to the 4-year-olds performance of our typically developing sample. This somewhat explains why I did not find significant difference between the nonverbal and the verbal FBT as it was expected. Based on the two language tests, which were administered these children language ability was already at a level when typically developing children pass verbal FBTs. However, it still does not explain the relatively poor performance on the nonverbal FBTs. Based on the level of language ability one would expect significantly better performance than chance level on both tests. Therefore subgroups were created within the DLI sample in order to find more homogeneous subgroups. Statistical method was used to create subgroups, a method applied in the SLI literature too. The results of the cluster analysis, after changing all variables to z-scores are presented in a dendogram in Figure 11. two subgroups were separated, the first group included children 4, 5, 7, 2, 3 and 1, while the second group included children 15, 21, 6, 9, 8, 16, 17, 18, 13, 19, 10, 12, 14, 20 an 11.

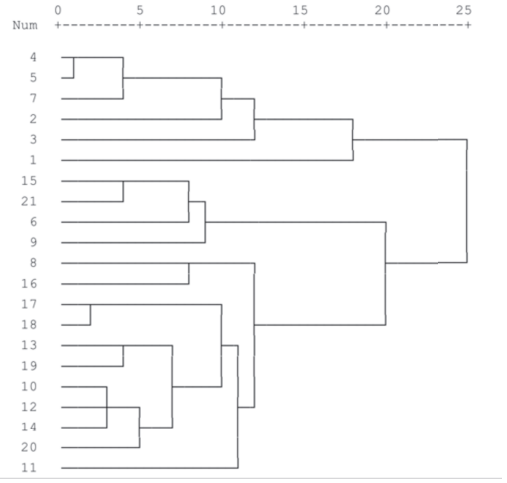


Figure 11. Dendogram of the results of the cluster analysis on DLI children's performance patterns.

One-way ANOVA revealed significant difference between the two groups in the performance of nonverbal FBT ($F(1, 19) = 37.22$, $p \leq 0.001$), the Peabody ($F(1, 19) = 13.26$, $p \leq 0.005$) and the TROG-H ($F(1, 19) = 22.43$, $p \leq 0.001$)[15] suggesting that the second group outperforms the first group in all of these tests.

The second group's (n=15) performance on both FBTs is above chance level (verbal FBT 73%, nonverbal FBT 60%), but not significantly (Binominal, $p>0.05$ for both tests) while the first groups' performance on both FBTs is well below chance level (verbal FBT 33%, nonverbal FBT 0%). Considering that the second group performed significantly better on both language tests, it could be a plausible suggestion that the two abilities, namely language and false belief understanding relate in this population. However the correlational data did not underpin this, since no correlation was found within the two subgroups.

As it can be seen in the dendogram group 2 can be divided into 2 further subgroups; subgroup 2a includes children 15, 21, 6 and 9, while subgroup 2b includes 8, 16, 17, 18, 13, 19, 10, 12, 14, 20 and 11. However group 2a includes only 4 children, which makes statistical comparisons difficult, thus I am going to concentrate on the first two groups created within the sample (group1 and group2) during discussion.

5.1.8. Conclusion

The validation of the nonverbal

The first major finding of *Study 1* was the similar developmental trajectory found between the verbal and nonverbal FBT. 3-year-olds did not pass the FBTs above chance level, while 4-and 5-year-olds passed both of them above chance, however only 5-year-olds passed it significantly above chance. In general, children's overall performance corresponds with the data of the false belief understanding literature, and suggests that the new nonverbal FBT is adequate in the critical age range of FB understanding.

The second major finding was that in spite of the similar developmental trajectory, there was a significant difference between the two FBTs, and subsequent analyses revealed that at the age of 3 the performance on the nonverbal FBT is better than on the verbal FBT, however no such difference was found at the age of 4-and 5 years. This suggests that the verbality of the FBT makes it more difficult for young children to pass the verbal FBT, but at an older age

[15] A more robust test, the Welch-test revealed the same significant differences between the two groups.

when probably the verbal abilities are better, this difference disappears. This is also suggested by the results, that language tests (PPVT and TROG-H) correlate with verbal FBT but not with the nonverbal FBT. These results will be further discussed later.

Third, although it was hypothesized that the two FBTs measure the same ability, no correlation was found between them after controlling for the effect of age. This result might raise questions about the validity of the nonverbal FBT, however I argue that this finding is due to the different routes the theory of mind ability was accessed, for a number of reasons. It was found that when the significantly more difficult Smarties task was not taken into account in the verbal FBT score (raw score), the correlation between the two FBTs reached tendency between the two tests, which suggests that the different false belief tasks in the two FBTs (verbal: location change + false identity, nonverbal: location change + true belief) at least partially responsible for the lack of correlation. Another argument comes from the literature; Kobayashi and colleagues (2007) also found significant difference in the performance of verbal and nonverbal FBT –however in adults – and found a difference in their neural correlates too.

A further argument is the result found when only those children were entered into the analyses who performed consistently on the two FTBs, thus excluding children with a "fragile" ToM. There has been a debate about the test-retest reliability of the FBT. Mayes et al (1996) found poor test-retest reliability in ToM tasks at this age range, while Hughes and her colleagues' (2000) results did not confirm this phenomena; they found good reliability for a bigger age range. Anyway, in the current study, after the exclusion, the relation between the two tasks was significant, which suggest that another partial reason of the lack of correlation was that children's performance on FBTs is not reliable at this age.

And finally, it is worth to discuss, how it is possible that Call and Tomasello (1999) did find a correlation between their version of nonverbal FBT and the verbal FBT but in this study with another version of nonverbal FBT I did not. Call and Tomasello applied Pearson's correlation and did not control for the effect of age (Pearson's correlation reveled significant correlation between the two FBTs in this study too). Thus, it is possible that what they found was the effect of age, or in other words the effect of development on the two tasks. The new nonverbal FBT clearly has advantages compared with Call and Tomasello's test. The test takes less time to administer, since no training phase is necessary, which considering young children's limited memory and attention capacity, is a very important feature. This could also be the reason why Call and Tomasello's test was passed only by 5-year-olds and not by 4-year-olds.

Another advantage of the test is that it is easier to administer since only one experimenter is needed.

In summary, these findings suggest that the new nonverbal FBT is a valid test of false belief understanding and therefore a valid test of an important aspect of theory of mind, which has advantages compared to other nonverbal FBTs.

Relations between false belief tests (nonverbal and verbal) and language tests

One of the aims of developing the nonverbal FBT was to get closer to the question regarding the developmental relations between language and ToM. It has been argued by many authors in the relevant literature (e.g. Astington & Jenkins, 1999 or de Villiers, 2007) that in order to get a clearer picture about this relation, a nonverbal ToM task is needed to avoid the possible effect of the verbality of ToM tasks, which are in most cases FBTs. Therefore, the results on this issue with the new nonverbal FBT are especially important. It was found that while the verbal FBT correlated with both language tests, the nonverbal FBT did not. This indicates that earlier findings between language and ToM are the by-products of the verbality of the FBT, and the relation found between the two abilities did not reflect a causal, essential relation. These results, however poses further questions about the nature of theory of mind ability. At least two alternative interpretations are possible. The first is called the *double manifestation hypothesis*; the two ways of testing (verbal and nonverbal) mobilize the ToM competence through two distinct channels. This hypothesis has a weaker and stronger version; the *weaker* assumes a unitary ToM competence, while the *stronger* assumes that the two distinct channels mobilize at least partially distinct ToM systems. The idea of a non-unitary ToM ability would not be completely genuine in the literature; there are some theories, which suggest this idea (e.g. Sperber, 2000; Happé & Loth, 2002; Tager-Flusber & Sullivan, 2000) (more about this issue in study 4).

The second possible interpretation is the *verbal filter hypothesis*, that the manifestation of a unitary theory of mind (false belief understanding) competence is constrained by linguistic abilities if it is accessed through a verbal channel, especially in young children and individuals who have language deficit. Although when ToM competence is accessed nonverbally, no such additional ability is needed, thus the nonverbal FBT reflects better the ToM competence (also suggested by Chandler et al, 1989 or Fodor, 1992) (certainly there is still a certain amount of performance limit, such as working-memory, inhibition, action

regulation etc.). The results of the study supports slightly better the second hypothesis rather than the double manifestation hypothesis, since it was found that the nonverbal FBT is less difficult for children at the age of 3. However the double manifestation hypothesis cannot be excluded as a possible interpretation either.

Language and FB understanding in developmental language impairment

The first important result of *Study 1* with this population is that children with DLI do have a delay in FB understanding and this resonates with other findings in the literature. It was found that children with DLI did not pass the nonverbal and the verbal FBT above chance around the age of 6.

Typically developing 3-year-olds performed significantly better on the nonverbal FBT than on the verbal FBT, suggesting that for children with lower language ability the nonverbal test is easier. Since earlier studies (Miller, 2001) on children with SLI found that by decreasing the linguistic demand of the FBT the children's performance increased, an even more robust difference was expected in DLI children between the 2 FBTs. However, no such difference was found with these children. Moreover, their performance on the nonverbal FBT, but not on the verbal FBT was below chance level, although the mean age of these children was 6;5 years. Since these unexpected results could have been caused by the heterogeneity of the DLI group, subgroups were created within this sample. After applying statistical method to create subgroups, 2 subgroups were found, one with a better performance on the nonverbal FBT and on both of the language tests than the other subgroup. Within this "advanced" subgroup, children still did not perform significantly above chance level on any of the FBTs, and within the other subgroup they performed below chance level on both FBTs. It is important to note, that no correlation was found within these two groups, and that neither the IQ nor the age was significantly different between the groups. It suggests that the "advanced group" has better nonverbal social abilities, or at least nonverbal ToM ability and better language ability too, but these abilities are independent from each other.

Another interesting finding of the study, which was already briefly mentioned, is the lack of correlations within the two subgroups. Unlike in the typically developing sample the two groups of children with DLI tested in the study did not show correlation between the verbal FBT and the language tests or between the two language tests measuring grammar and vocabulary. The most probable reason of this is that these groups' language deficit is still

heterogeneous, and maybe at the level of individuals we would get very different patterns; for instance different patter in DLI children's grammatical ability and vocabulary.

In sum, although it was expected, the nonverbal FBT was not easier then the verbal FBT for children. Since no relation was found between their performance on FBTs and language ability, it is improbable that this result would reflect a strong relation between FB understanding –would it be measured with a verbal or nonverbal test – and language ability in this population. However, our hypotheses were based on research with the SLI population, and our sample was a DLI sample, which is inevitably more heterogeneous and this could cause our different results.

5.2. Study 2: Shorten version of the nonverbal false belief test

5.2.1. Background

The development of the shorten version of the nonverbal FBT (mini nonverbal FBT) was twofold. First, make the tuning phase even shorter to decrease the recourse of other cognitive abilities, such as attention, memory etc. Our experiences with the original, longer version of the nonverbal FBT suggested that children are able to get tuned to the task situation, in which the experimenter started a story and they had to finish it, before the end of the tuning phase. The second aim was to avoid the criticism that both the nonverbal and the verbal (!) FBT could have been solved by using lower level processes, e.g. by associating the first character with his/her "place". E.g. in the Sally Anne situation the child led the character to the place where he/she put the ball, and not where it really was, because that is his box/basket or that was the place where she/he was seen to manipulate the ball. It is worth to note though, that the vast majority of – certainly verbal – false belief tests do not include any other tasks – true belief tasks for instance – which would control this possibility, which means that we could drop out the ~90% of the false belief understanding literature and therefore a big amount of theory of mind literate too. Another data that suggests the validity of the new nonverbal test is the connection found between the novel nonverbal FBT and one of the ADOS indexes, at children with autism (Győri et al, 2007). It strongly suggests that our novel nonverbal FBT mirrors social capacities-skills and not simple associations.

Because of the first aim the tuning phase was shortened, leaving only the very first task, the "carrying blocks" and the "demo" to tune the children to use attribution of mental states as a task solving strategy.

Our second modification was motivated to avoid a possible but not probable non-mentalistic task solving strategy, as described above, namely using simple association between the first character and his/her "place". Another true belief trial was added to the nonverbal and verbal FBTs, which contained a visible location change. The first character placed the object into one of two places, but did not leave. Therefore she/he was able to see that the second character replaced the object to the second place. In this true belief condition, the first character knows where the object is, but the location is different from the place where she/he originally put it.

Similarly, not only the nonverbal FBT but also the verbal FBT was changed to make the structure of the two test the same. Therefore the verbal FBT also contained two location change false belief and two true belief trials; one *without l*ocation change and one *with* location change that is watched by the first character (see Figure 12.).

Tuning Phase					
1.	„Carrying cubes"	1 puppet	Repetitive action	Pass:→ 2	Fail:→ End
D	"DEMO" FALSE BELIEF „garages"	2 puppets	False belief situation, but the researcher acts it out	→ T1	
Test Phase (random order)					
T1	FALSE BELIEF „Sally-Anne"	2 puppets	False belief situation, with location change		
T2	TRUE BELIEF „Cooking"	2 puppets	False belief situation, but without location change		
T3	TRUE BELIEF „Flowers"	2 puppets	False belief situation, but the first character watches the location change		
T4	FALSE BELIEF „Sleeping"	2 puppets	False belief situation, with location change		

Figure 12. The short nonverbal false belief test.

5.2.2. Hypotheses & questions

Study 2 was designed to test the following hypotheses and questions:

(1) whether children pass the short nonverbal FBT earlier than the verbal FBT, around the age of 3.

(2) similar relations are expected between the language tests and the two FBT like in Study 1; namely significant correlation between the verbal FBT and both the Peabody and the TROG-H

5.2.3. Participants

The participants were 39 typically developing children, 21 3-year-old (9 boys and 12 girls) and 18 4-year-old children (6 boys and 12 girls). Since *Study 1* showed that our nonverbal FBT is sensitive of the developmental shift at the age of 4 found in theory of mind development, in this study we focused on the two age groups of three and four years olds and no 5-year-olds were tested. All children were native Hungarian and were recruited from a kindergarten in Budapest (Gróf Brunszvik Teréz Napköziotthonos Óvoda).

5.2.4. Materials

Short nonverbal false belief test: described above, see Figure 14. Just like in *Study 1* both raw scores (0-4) and pass/fail scores were used in the analyses. The pass/fail score was very strict, since the child had to pass all four trials in order to pass the mini nonverbal FBT.

Verbal false belief test: the test contained four test trials. The two false belief trials and the true belief trial without location change were the same as in *Study1*. The new true belief trial with location change was the following: the first character appeared on the scene putting an object into one of two hiding places. Then the first character does not leave the scene, but stays there and watches what the other character does. We also stress this when we verbally tell the story to the children (see Appendix 2). The second character then goes to the same hiding place, takes the object out and puts it into the other hiding place. The second character leaves. Then we ask the children the usual questions, just like in *Study 1*; where the first character will look for the object, plus the memory, reality and identity questions. Children can pass this trial if they realize that since the first character has seen where the second character replaced the object, the first character knows where the object really is, that is if they attribute true belief to the first character (and certainly if they pass the memory, reality and identity questions too). Just like in *Study 1*, we worked with both the raw scores (0-4) and pass/ fail score (0/1 and not evaluable). The exact same pass/fail criteria were used as for the mini nonverbal FBT and it was calculated the same way, using the strict 4/4 criteria.

Peabody and TROG-H: both described in *Study 1*. The scores were calculated the same way as in *Study 1*.

5.2.5. Results

Raw scores

Figure 15 presents the percentage of correct trials in the verbal and nonverbal FBT as a function of age. Overall, *3-year-old children* (n=21) failed to select the correct hiding place above chance in the verbal FBT (ns), but Binominal test revealed, that children did select the correct hiding place in the nonverbal FBT above chance level, but it did not reach significance (Binominal, $p>0.05$). Only one child led the puppet to correct box on 3 out of 4 trials (None selected it on all of the trials) in the verbal FBT. In contrast, in the nonverbal FBT 3 children selected the correct place on all of the trials and 10 selected it on 3 out of 4 trials.
Overall, *4-year-old children* (n=18) selected the correct hiding place greater than chance level in both the verbal FBT and the nonverbal FBT, but again it did not reach significance (Binominal, $p>0.05$). In this age group, ten of the children selected the correct place on 3 out of 4 trials in the verbal FBT, and 12 of the children selected the correct hiding place in the nonverbal FBT. (No children selected the correct place on all of the trials).

A 2 x 2 (age x FBT) ANOVA with age (3-and 4-year-olds) as a between subject factor and FBT as a within-subject factor revealed a significant main effect for FBT $F(1, 37) = 23.43$, $p \leq 0.001$. Regardless of age, children's performance on the nonverbal FBT is better compared with the verbal FBT. The main effect of between-subject factor age ($F(1, 37) = 16.07$, $p \leq 0.001$) was also significant, indicating that overall, 4-year-olds performance on the FBTs was better than the 3-year-olds performance. In addition there was a significant effect of interaction $F(1, 37) = 15.1$, $p \leq 0.001$. Subsequent analyses comparing the mean score of the nonverbal FBT and the verbal FBT within each individual age group revealed that the performance on the two tests was significantly different at the age of 3 years ($t(20) = -5.775$, $p \leq 0.001$) but not at 4 years (4-year-olds: $t(17) = -0.77$, $p > 0.05$), suggesting that 3-year-olds performance on the Nonverbal FBT is significantly better than on the Verbal FBT [16]

[16] Just like in Study 1, considering that the variables of the FBTs are not typical continuous variables (0-4) a more robust nonparametric analysis was also conducted, and the same significant results were found. The Mann-Whitney test revealed significant difference between the 3-and 4-year-olds in both the verbal ($U=43.5$, $p<0.001$) but not in the nonverbal FBT. Further comparisons found significant difference at the age of 3 years between the verbal and nonverbal FBTs (Wilcoxon test: $z= -3.6$, $p > 0.001$, the mean of ranks in favor of nonverbal FBT was 9.82, while the mean of the ranks in favor of verbal FBT was 4) but not in the 4-year-olds ($z=-0.775$, $p > 0.05$,

Figure 13. Mean percentage of performance on the verbal and on the nonverbal FBT as a function of age.

A possible reason of 3-year-olds relatively good performance on the nonverbal FBT is, that older 3-year-olds performed very high in the task (full of either 3 or 4 correct answers), while younger 3-year-olds performance was still poor[17]. Such difference could suggest that these children could pass the nonverbal FBT only a few months earlier than the verbal FBT. To test this hypothesis the group of 3-year-olds was divided into two groups; *younger 3-year-olds*: 3;0-3;5, n = 9, and *older 3-year-olds*: 3;6-3;11, n = 12. But the subsequent analyses did not find any significant difference between young 3-year-olds performance (3-3;5 years) and older 3-year-olds performance (3;6-3;11 years) in the nonverbal FBT ($F(1, 19)=0.23$, $p > 0.05$), suggesting that the young 3-year-olds already performed at a relatively high level on the nonverbal FBT (see Figure 14).

the mean of ranks in favor of nonverbal FBT was 6.86, while the mean of the ranks in favor of verbal FBT was 6).

[17] Note that it still would not explain why older 3-year-olds did not perform better on the verbal FBT.

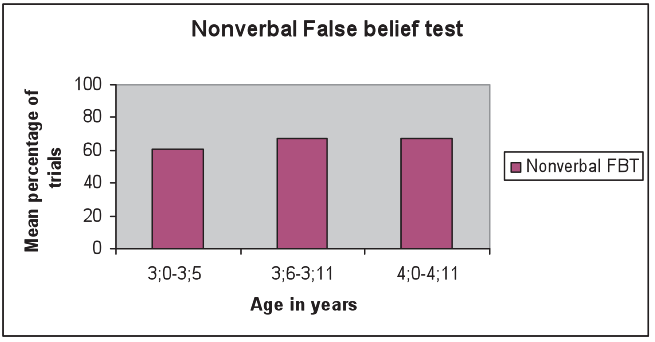


Figure 14. Mean percentage of performance of the nonverbal false belief test as a function of age.

Pass/Fail scores

Again, using strict criteria, children passed both Nonverbal and Verbal FBT if they passed all of the four trials. The results are very surprising with this score. At the 3-year-olds, none of the children passed the verbal FBT and only 3 passed the nonverbal FBT. At the four-year-olds no child passed the verbal FBT and the nonverbal FBT.

Further analyses were needed to find out whether one of the trials in the FBT is responsible of these extremely poor performances. We found that both in the verbal and in the nonverbal FBT the performance on the new true belief trial with location change was under chance level (Figure 15); 14 children (~36%) passed it in the nonverbal FBT and 11 children (~28%) passed it in the verbal FBT. McNemar test, however, did not reveal significant difference between the location change TB trial and the two FB trials either in the verbal and or in the nonverbal FBT (verbal FB trial1, $p > 0.05$; verbal FB trial2, $p > 0.05$; nonverbal FBT1, $p > 0.05$; nonverbal FBT2, $p > 0.05$). However, the difference was significant between the two true belief trials: with location change and without location change (McNemar verbal TB no location change trial $p \leq 0.05$, nonverbal TB no location change trial $p \leq 0.001$).

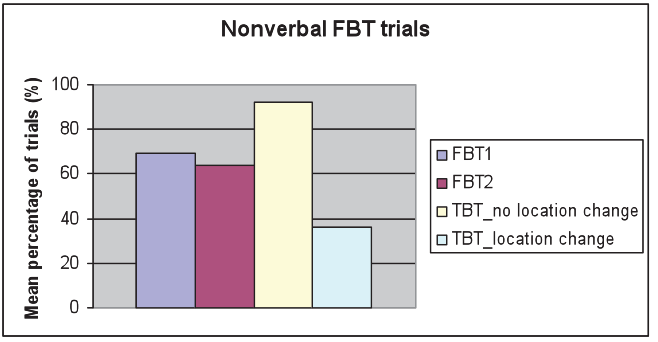


Figure 15. Mean percentage of trials of the nonverbal FBT.

Further analyses found no difference regarding the order of the presentation of the trials. Thus even when the TB trial was presented first, the children's performance was not better on the TB trial, than when it was presented after a false belief trial. Therefore this phenomenon cannot be explained as when children were presented with the FB first they used the same strategy to solve the TB trial and that is why their performance was low. Also, since we found the same pattern of performance with the verbal and the nonverbal FBT it is very unlikely that in nonverbal FBT children would copy some simple, non-mentalistic strategy from the demonstration. Interestingly, however when the TB trial was compared between the two age groups, it was found that 3-year-olds performance was better both on the verbal and on the nonverbal TB trial (Figure 16). Fifty-eight percent of the 3-year-olds passed the verbal TB trial and 48% the nonverbal TB trial. In contrast, only 28% of the 4-year-olds passed the verbal FBT and 33.3% the nonverbal TB trial. The difference is not significant between the two age groups either in the verbal (Fisher's test, $p>0.05$) or in the nonverbal (Fisher's test, $p>0.05$) TB trials.

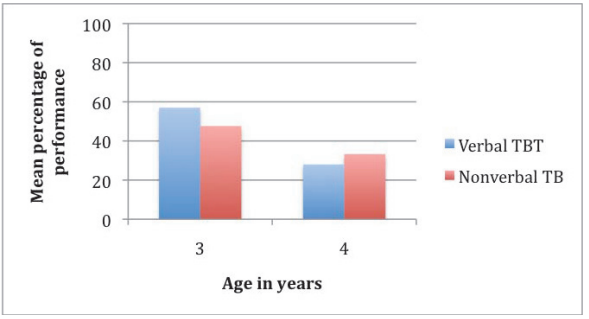


Figure 16. Children's performance on the verbal and the nonverbal TB trial.

These results suggest, that while children's performance on the FB trials increases, as children get closer to their fourth birthday, their performance on the TB trials decreases.

Relationship between language and false belief understanding

Raw scores were used in the partial correlation, controlling for age to analyze the relations between the language tests; Peabody and TROG-H - in the TROG the number of correct blocks was used – and the false belief understanding tests; Verbal FBT (0-4) and Nonverbal FBT (0-4). Table 5 presents the correlational results. Here, again it was found that both the PPVT and the TROG-H correlated with the Verbal FBT (PPVT $r = 0.336$ $p \leq 0.05$; TROG-H $r = 0.316$ $p \leq 0.05$), but not with the Nonverbal FBT. The two false belief tests do not correlate with each other either. Certainly the language tests strongly correlate with each other $r = 0.679$ $p \leq 0.001$).

	Verbal FBT	**PPVT**	**TROG-H**
Nonverbal FBT	n.s	n.s.	n.s.
Verbal FBT		**0.336 ($p \leq 0.05$)**	**0.316 ($p \leq 0.05$)**
PPVT			**0.679 ($p \leq 0.001$)**

Table 5. Partial correlation between the language tests and the false belief understanding tests, controlling for age.

5.2.6. Conclusion

The aim of this study was to twofold. First, by shortening the nonverbal FBT to reduce young children's cognitive load presented by the task and by doing this to potentially increase the number of young children – around the age three – who are able to pass the nonverbal FBT. The results supported our hypotheses, since 3-year-olds, both younger and older 3-year-olds performed significantly better on the nonverbal FBT than on the verbal FBT (Note that this pattern of results was not found in Study1). However, the correlational results were the same as in Study1, suggesting that only the verbal FBT has a developmental relation with language, but the nonverbal FBT does not.

The second aim of the study was to exclude a possible nonmentalistic solution of the nonverbal FBT; namely that kids can pass it by simply looking for the object at the hiding place, which was earlier associated with the first character. To avoid such a criticism a location change TBT was included both in the verbal and the nonverbal FBT. The results were unexpected; children performed somewhat worse, but not significantly worse on the location change TBT than on the FBT. Therefore when used a conservative pass/fail score the 4-year-olds did not pass either the verbal FBT or the nonverbal FBT above chance (and certainly neither did the 3-year-olds). A major question of Study 2 thus, is why children's performance was so low on the location change TBT.

The first suggestion could be that these children solved the nonverbal FBT based on some associative strategy therefore it does not measure false belief understanding. However, we found very similar data with verbal FBT too, so these results would suggest that the verbal FBT, which is the basis of the theory of mind literature does not measure FBT either. Also, however the performance was lower than the performance on the other three trials, the difference was not significant.

The possibility that in the nonverbal FBT the demo would have suggested a task solving strategy other than false belief understanding, e.g. lead the puppet to the empty place is not probable since, again, similar results were found with the verbal FBT in which no demo is included.

Another possibility that these children had difficulty with switching between false and true belief understanding, however, it would have caused an order effect; both true and false belief tests have better performance if they are presented first. After testing this possibility the results did not support this idea.

A simple explanation could be that children did not know somehow that the presence of the puppet meant that he/she saw and therefore knew what the other character did. This could be the case with the nonverbal FBT, but in the verbal FBT the experimenter explicitly states that the first character sees what the second character does. However, since this was not explicitly controlled in any of the tests – for instance the experimenter did not ask the children whether the first character saw that the second character replaced the object, this explanation cannot be excluded.

Although the data with the location change TBT is surprising, is not without precedent in the ToM literature. As we saw it in Chapter 2, a systematic comparison of different TBTs (Lohmann et al, 2005) revealed that 3.5-4-year-old children's performance on the location change TBT is just above chance (not significantly above though), while children's performance on the no location change TB trial is much better. Our results are in line with these findings, since the 3-year-olds in this study also performed around chance, while 3-year-olds performance on the no location change trial was well above chance. Unfortunately Lohmann, Carpenter and Call (2005) do not report any data above the age of 4. Another study in the literature, which is in line with our results, is the Roth & Leslie's paper (1998), as we saw it in Chapter 2. The authors found the same phenomena, however using different TB tasks; young 3-year-olds' performance was better – not significantly better – than older 3-year-olds' (3;9-4;0) performance. These results suggest that around the age of 4, when children begin to understand false beliefs they will have an overall preference towards false belief situations. It could be because they find the TBT less interesting and that is why they do not pay as much attention to it, or that they enjoy the trickery part of the FBT and they use it in situation too, where it is not adequate.

5.3. Study 3: Testing the complement hypothesis on typically developing children and children with autism

5.3.1. Background

I discussed in the introduction a very radical statement regarding the relation of theory of mind or in the hypothesis later version FB understanding and language acquisition. De Villiers claims that the mastery of early sentential complements is the prerequisites of passing FBT (most recently, de Villiers, 2007) when a collection of FBTs and language tasks were administered 3 different times within about a half a year period. The FBTs were an unexpected contents task (Perner, Leekam, & Wimmer, 1987) (or Smarties task), an unseen displacement (Wimmer & Perner, 1983)(or Sally-Anne task) and an explanation of action task (Bartsch & Wellman, 1989), children could receive maximum 2 points for each task and, a child was credited with passing the FBTs if he/she received at least 5 points out of 6. The language tasks were Memory for complements (criterion of passing is 10/12), different scores of spontaneous speech (MLU, total IPSYN score and sub-scores of IPSYN) and the Complements in wh-questions, which turned out to be too difficult for children at this age. The study and the hypothesis have been criticized in many ways, the aim of this study is to focus on some of these. First of all, one of the major critics of the research of the developmental relation between language and ToM is that all of these data were gathered with a verbal FBT, which leaves open the possibility that the relation found is only due to the verbality of the test and not to the relation of the abilities (e.g. Astington & Jenkins, 1999). In Study1 I found that the nonverbal FBT did not correlate with the language tests, and it raises the question whether I will find any relation between the nonverbal FBT and the complement test as de Villiers hypothesis would suggest.

A remarkable portion of the critics focuses on the tests that were used in the de Villiers paper (de Villiers & Pyers, 2002). FB understanding was measured by 3 FBTs one of which was an explanation of action: after presenting a unexpected content task the child was asked why the first character was looking in the empty box? Similarly, the child passed the unseen displacement task if she/he could give an adequate answer why the first character looked for its object in the empty box. Most FB study does not include these questions as a criterion of passing it, since they make the task unnecessarily difficult (also discussed in Chapter 2). Not surprisingly, these tasks were more difficult than the simple unexpected content task, and this

could lead to result of children passing FBTs later than Complement tasks. As it was also mentioned earlier, the study was criticized of using spontaneous speech as a measure of language ability (e.g. Slade & Ruffman, 2005) instead of using standardized tests, which represent the comprehension of different language aspects. Thus in the current study two unseen displacement and one unexpected content tasks were administered as the FBTs and the TROG-H (receptive grammar) and PPVT (receptive vocabulary) were used as language tasks in addition to the Memory for complement task.
And finally, a third field of critics come from studies conducted in languages other than English. These studies suggest the de Villiers hypothesis is not universal across languages (Tardif & Wellman, 2000; Perner et al, 2003; Cheung et al, 2004; Tardif et al, 2007), thus it is possible that I will not find the same relation as de Villiers did with the verbal FBT either.

As we saw it in part 4.2.2.4.1. de Villiers hypothesis was tested on children with ASD too. In earlier studies (e.g. Happé , 1994) a close relation was found between language ability and FB understanding in individuals with ASD (or at least in those who have at least some language and pass FBT). Thus, if de Villiers' hypothesis is true, we can expect at least the same or even stronger relation between sentential complements and FBT. Both in the Tager-Flusberg (2000) and in the Lind & Bowler (2009) studies a strong relation was found between complements and FB understanding with verbal FBT.
Krisztina Stefanik (2005) using the new nonverbal FBT introduced in Study 1, found that "relatively higher functioning"[18] children with ASD performed somewhat better on the verbal FBT than on the nonverbal FBT, however the difference was not significant. Another interesting result of the Krisztina Stefanik's PhD Dissertation (2005) regarding the nonverbal FBT is that it correlated with the ASD children's social/communicative behavior (measured by ADOS), while the verbal FBT did not show such a relation (also in Győri et al, 2007). According to the authors these findings suggest a closer relation between the everday social behavior and the nonverbal FBT but not with the verbal FBT. These findings also bring on the possibility of verbal compensation, which will be further discussed after the current study.

5.3.2. Hypotheses and questions

Study 3 has the following questions and hypotheses for *typically developing children*:

[18] The ASD sample was not strictly high-functioning children with ASD, but their nonverbal IQ (Leiter I.S) was above 60.

(1) whether the de Villiers data will be replicated when false belief understanding is measured by a nonverbal test
(2) whether I find the same predictive relation between the different language tests and verbal FB tests
(3) whether de Villiers result will be replicated in Hungarian language too

Hypotheses for *children with ASD*
(1) children with ASD show the same or even stronger relation between the verbal FBT and the language tasks including the complement task in Hungarian language too
(2) the relation will not be significant if FB understanding is accessed with the nonverbal FBT

5.3.3. Participants

Two groups of children participated in our study; typically developing children and children with autism spectrum disorder, all children were native Hungarian speakers.
Thirty-four *typically developing children*, twelve 3-year-olds (mean age: 3;7 years, SD: 3.28 months, 8 girls and 4 boys), thirteen 4-year-olds (mean age: 4;6 years, SD: 3.11 months, 8 girls and 5 boys) and ten five-year-olds (mean age: 5;10 years, SD: 6.19 months, 5 girls and 5 boys) participated in the study. All of them were recruited from public preschools, kindergartens (Győri Apor Vilmos Katolikus Iskolakozpont, Veres Peter).
Sixteen *children with autism spectrum disorder* (1 girl) aged 7;7 to 11;9 mean age 10;2, SD: 15.17 months). All children were recruited through the Autism Foundation, Budapest, Hungary. Diagnosis of autism was made on the basis of DSM-IV criteria (APA, 1994). The children's IQ scores were obtained using the Hungarian version of the Wechsler Intelligence Scale Revised (WISC-R, Kaufman, 1979, in Hungarian: Mawgyi-R), the mean IQ was 79.7 (SD: 14.5) (verbal IQ: 81.3, SD: 13.5; performance IQ: 82.6, SD: 17.4). The main criterion of selecting children was the verbal mental age based on the investigation of Happé (1995) according to which children with autism do not tend to pass the (verbal) false belief tests under the verbal mental age of seven, but they reach ceiling effect above the verbal mental age of thirteen. The ASD group was identical to the one tested in Stefanik's Dissertation (2005).

5.3.4. Materials and procedure

Language

Language level was assessed by the *Peabody* Picture Vocabulary Test (PPVT; Csányi, 1974) which is a measure of one-word receptive vocabulary and with the Hungarian version of TROG (Bishop, 1983), *TROG-H* which is under standardization and which measures the receptive grammar. In the lack of a complete Hungarian standard we used the raw scores at both tests; the number of correct answers in the Peabody and the number of correct blocks in the TROG-H.

The vocabulary of children with ASD was measured with the *'vocabulary subtest'* (*Maw-voc*) of the Hungarian version of Wechsler Intelligence Scale Revised (Mawgyi-R).

Complement tasks: as the age of the two groups of participant was significantly different and therefore their linguistic abilities too we applied two tasks which differed in their difficulty to test sentential complements. *Memory for Complements* was used to test typically developing children. An example: The girl said to her sister that she brought some apples, but she really brought some oranges. "What did the girl say?" The correct answer was "that she brought some apples" ("apples" was accepted as correct answer too). We presented 16 sentences and children passed the test if they answered minimum 14 questions correctly
(Criteria based on De Villiers & Pyers, 2002).

Complements in wh-questions were used to test children with autism. An example: A boy was having chocolate in the school at noon. Later he went home and played with his toys. That evening he said to his mum "I ate chocolate this noon!" We asked then two questions: (1)When did the boy say what he ate? The correct answer was "that evening". (2) When did he say he ate? The correct answer was "that noon". 8 stories were presented; criterion for passing was set at 13 or more out of 16 (as two questions were given after every story). (The entire test with pictures and the text in Hungarian can be found in Appendix 3)

Theory of Mind

Two different types of false belief tests were administered; a verbal and a nonverbal test. In the verbal test three standard false belief tasks were presented; two location-change false belief tasks based on Baron-Cohen, (1985) and an unexpected-contents false belief task based on Perner, Leekam, and Wimmer (1987). Children passed the *verbal FBT* if minimum 2 tasks were passed (and all the control questions were answered correctly). The *nonverbal FBT* was

the same as in *Study 1*, it contains two false belief tasks and a true belief task. Children passed the nonverbal FBT if all the three tasks were passed.

5.3.5. Results

5.3.5.1. Results on typically developing children

The mean scores and standard deviations in brackets of the Peabody (PPVT), TROG-H, Memory for complements task, nonverbal FBT and the verbal FBT within the 3-year-olds, the 4-year-olds and the 5-year-olds are displayed in Table 6.

	Peabody	TROG-H	Memory for Complements	Verbal FBT	Nonverbal FBT
3-year-olds	44.25 (11.98)	4.7 (2.5)	12.17 (3.81)	1 (0.74)	2 (1.04)
4-year-olds	49.5 (14.32)	7 (3.82)	11.08 (3.85)	1.67 (1.07)	2.6 (0.89)
5-year-olds	69.9 (7.31)	15.25 (1.48)	14.8 (1.23)	2.6 (0.516)	2.9 (0.32)

Table 6. Children's mean (SD) performance on the different tests as a function of age.

As presented in Figure 17, 3-year-old children did not pass the Memory for Complements task above chance level, four-year-olds' performance was 64% and the five-year-olds' performance was already 92%. From the three 3-year-old children who passed Memory for Complements task, two passed the verbal FBT and one passed the nonverbal FBT. There was one child who passed the verbal FBT but did not perform on the Memory for Complements task above chance and four passed the nonverbal FBT without performing on the Memory for Complements task above chance. Among the 4-year-olds seven children performed above chance on the Memory for Complements task and 6 of them passed the verbal FBT and seven passed the nonverbal FBT. However an additional one child passed the verbal FBT and three children the nonverbal FBT who did not perform on the Memory for Complements task above chance. And finally, all of the 5-year-olds performed on the Memory for Complements task above chance and passed both the verbal and nonverbal FBT. Clearly, the children who passed either the verbal or the nonverbal FBT, in a total of seven children, are challenges for

the Complement hypothesis. Note that 2 of these children passed both the verbal FBT and the nonverbal FBT but still had a low performance on the Memory for Complements task. In contrast, in the de Villiers study there were 5 children, who did not fit into the expected pattern, namely failed on the complement task but passed the FBT. De Villiers argues that all of these children produced at least one perfectly formed complement in spontaneous speech, thus gave evidence of "productive command of complementation" (de Villiers & Pyers, 2002, p. 1051).

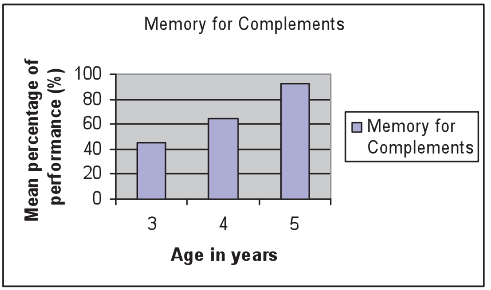


Figure 17. Children's performance on the Memory for complements task

I compared the Hungarian children's performance to the children's performance in de Villiers' study conservative criteria of passing the task – minimum of 14 out of 16 –, no three-year-old (0%), only 3 four-year-old (25%) and 10 thus all of the five-year-olds (100%) passed the task. It differs from the data found by the author in native English speaking children, who passed the task around the age three (de Villiers, 1999).

The verbal and nonverbal FBT showed similar developmental pattern to the one we saw in Study 1, thus 3-year-olds performance was under chance level (verbal FBT: 25%, nonverbal FBT 41.67%,) 4-year-olds passed the test above chance level, but not significantly above chance (Binominal, $p>0.05$) (verbal FBT: 58.33%, nonverbal 63.33%) and all of the five-year-olds passed both the verbal and nonverbal FBT (Binominal, $p\leq0.01$). Finally, as mentioned earlier I found a significant correlation between the verbal and nonverbal FBT, even after controlling for the effect of age. This correlation further supports the validation of the nonverbal FBT, though raises the question why I did not find it in Study1. The question will be discussed in the Conclusion part.

In order to test the relation between the language tests and the FBT, partial correlations were computed using the raw cores of the tests. Results are presented in Table 7, the correlations in bold are significant as shown. After controlling for age, significant correlations were found between the verbal FBT and Memory for complements (r=0.36; $p \leq 0.05$) the verbal FBT and TROG-H (r=0.59; $p \leq 0.001$) the verbal FBT and Peabody (r=0.61; $p \leq 0.001$) the verbal FBT and the nonverbal FBT (r=0.37; $p \leq 0.05$), the Peabody and TROG-H (r=0.55; $p \leq 0.005$) the Peabody and Memory for complements (r=0.41; $p \leq 0.05$), and the TROG-H and Memory for complements (r=0.48; $p \leq 0.05$).

	Peabody	**TROG-H**	**Memory for Complements**	**Verbal FBT**	**Nonverbal FBT**
Peabody	1.0000				
TROG-H	**0.55(**)**	1.0000			
Memory for Complements	**0.41(*)**	**0.48(*)**	1.0000		
Verbal FBT	**0.61(**)**	**0.59(**)**	**0.37(*)**	1.0000	
Nonverbal FBT	0.3	0.118	0.089	**0.37(*)**	1.0000

Table 7. Partial correlations in typically developing children, controlling for age. (* $p \leq 0.05$, ** $p \leq 0.01$).

Interestingly thus, we found that the language tests including the memory for complements task correlated with the verbal FBT but not with the nonverbal FBT. However the two FBT, verbal and nonverbal did correlate with each other. However these are very interesting data themselves, to further investigate the question regarding the relation between language and false belief understanding, the scores of the Peabody test was also controlled. It has been found that the scores of the Peabody showed a strong correlation with general verbal ability in typically developing children (r=0.91, Dunn & Dunn, 1997[19]). By controlling for the effect of the Peabody I was able to check if the correlations found between the verbal FBT and the TROG-H or between the verbal FBT and the Memory for complements were only due to the general effect of language or a more specific relation can be found with one of these aspects

[19] Note however, that these data coming from a native English speaker population, and not Hungarian population.

of language. I found that the correlation between the verbal FBT and TROG-H still remained, which suggests the special importance of syntax – but not a specific aspect of syntax, e.g. complements – in false belief understanding. This result strongly resonates with Astington & Jenkins's (1999) findings.

The development of the Complements in wh-question task on typically developing children (CIW)

Following de Villiers (2002) and Tager-Flusberg (2000) method described above.

Since there is no data when Hungarian typically developing children pass the CIW task, or in more general how it develops during Hungarian language acquisition, 37 typically developing children were tested on the task. Thirteen 3-year-olds (mean age: 42.38 months; SD: 3.86 months), thirteen 4-year-olds (mean age: 53.54 months; SD: 3.07 months) and eleven 5- 6.5-year-olds (mean age: 71.36 months; SD: 6.2 months). The performance on the CIW task became as follows: No difference was found between the performance on the 2 different questions (the one who knew the correct answer to one of the questions, knew the correct answer to the other question too) therefore data were collapsed in further analyses. The 3- and 4-year-olds scored under chance level (3-year-olds' mean: 3.92; SD: 3.75 and the 4-year-olds' mean: 4.54 ; SD: 3.36), however there was a increase at the 5-6.5-year-olds' performance, which was well above chance level (mean: 11.55; SD: 3.53) as Figure 20 also presents. Thus, we found that native Hungarian children pass the CIW task approximately at the same age as native English children (Figure 18).

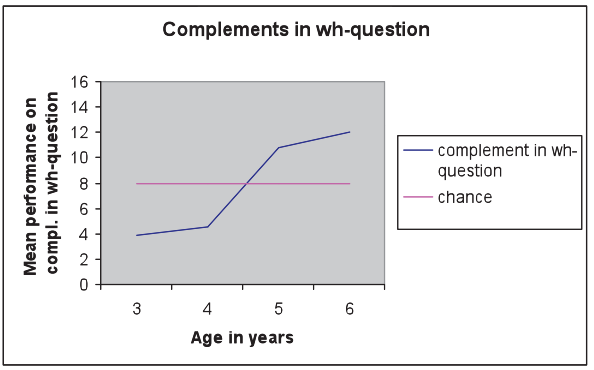


Figure 18. Children's performance on the complements in wh-question as a function of age.

5.3.5.2. Results on children with autism

Table 8 presents the means – standard deviations in brackets – of the tests applied in the study.

	MAW-VOC	TROG-H	Compl. in wh-	Verbal FBT	Nonverbal FBT
ASD children	17.25 (9.2)	14 (3.5)	9.25 (6.03)	2.35 (1.16)	2 (0.99)

Table 8. Chilren with ASD's mean performance on the tests, SD in brackets.

Children with ASD performed on the CIW task as the following: 11 children (68.75%) passed the test above 50%. Ten of these 11 children passed the verbal FBT – with the maximum raw score of three – and five of them passed the nonverbal FBT (with the score of three). However, there was one child who did pass the verbal FBT but his score was only 5 on the CIW task, also two children did pass the nonverbal FBT but received 5 (the same child who passed the verbal FBT too) and 0 scores on the CIW task. If Tager-Flusberg's criteria of passing the CIW task were applied (min. 13 out of 16), seven children (43.75%) passed the task. All of these children passed the verbal FBT and only three passed the nonverbal FBT. Therefore, four children did pass the verbal FBT but not the CIW task (their scores are: 5, 10, 12, 12) and four children did pass the nonverbal FBT but not the CIW task (their scores are: 0, 5, 12, 12).

Table 9 presents the results of Pearson correlation; the correlations in bold are significant, thus between the verbal FBT and all of the language tests: receptive vocabulary test (Maw-voc), the receptive grammar test (TROG-H) and the complement test (CIW), in contrast, no correlation was found with nonverbal FBT.

	TROG-H	Maw-voc	Compl. in wh-questions (CIW)	Verbal FBT	Nonverbal FBT
TROG-H	1.0000				
Maw-voc	**0.59(*)**	1.0000			
Compl. in wh-questions (CIW)	**0.68(**)**	**0.76 (**)**	1.0000		
Verbal FBT	**0.62(*)**	**0.67 (**)**	**0.82 (**)**	1.0000	
Nonverbal FBT	0.04	0.38	0.39	0.43	1.0000

Table 9. Pearson's correlations in children with autism. (* $p \leq 0.05$; ** $p \leq 0.01$)

It is clear that children with ASD show a strong association between their language skills and their verbal false belief understanding, as it was hypothesized and already found in earlier studies. This strong relation appears both in the strong correlations found between the verbal FBT and all of the language tests. However, this relation, just like in the typically developing children was not found with the nonverbal FBT as it was also hypothesized based on earlier research (Happé , 1994; Senju, in press).

In order to further investigate the relation between language and false belief understanding and to find more specific connection, partial correlations were applied to control for the effect of Maw-voc. This revealed that most of the correlations disappeared, except for the one between the verbal FBT and CIW, which was still strong ($r = 0.631$; $p \leq 0.05$). Nevertheless, when the CIW variable was controlled for, no correlation was found between any of the variables. These results are very interesting and support Tager-Flusberg's data (2000).

Eleven (68%) children passed the verbal FBT – it is not significantly above chance, Binominal, $p>0.05$ –, but the nonverbal FBT was passed only by 7 (43%) children, however the difference between the performance on the two tests is not significant (McNemar test, $p>0.05$) (Győri et al, 2007; Stefanik, 2005). There were 5 (31.25%) children who passed the verbal FBT but did not pass nonverbal FBT, in contrast to the only one (6.25%) child who passed the nonverbal but did not pass the verbal FBT.

5.3.6. Conclusion

Typically developing children

The first interesting result of the study is the correlation found between the verbal FBT and the nonverbal FBT in typically developing children. However the sample size was smaller than in Study 1 – where the correlation was only tendency –, but I argue that it strengthens the validity of the nonverbal FBT. As I mentioned in Study1, children around the age of 3-4 show a certain inconsistency in passing or failing FBT. Thus, I chose children who either failed or passed all of the trials within the verbal and nonverbal FBT, and these consistent children did show a correlation between the two tests. Following this line of thought, I checked whether children in this study presented a better consistency within each FBT. I found that 68% of the children gave consistent answers for both FBTs in the current study compared to the 56% consistency found in Study1. This difference in the consistency might be at least in part responsible for the different correlations found in the two studies regarding the FBTs. Another reason of this finding can be that while in Study1 a significant difference was found in 3-year-olds performance on the two FBTs, no such difference was found in this study (McNemar, $p>0.05$), although 3-year-olds still performed under chance level and their performances on the nonverbal FBT was somewhat better, but not significantly better.

The aim of this study, however, was to test de Villiers hypothesis – that the mastery of the memory for complements task is a prerequisite of false belief understanding – in Hungarian language and more importantly to test if the hypothesis is still hold when false belief understanding is measured with a nonverbal test. It was found that Hungarian children passed the Memory for complements task around the age of 4, when their performance was compared to chance level, but when de Villiers' stick criteria was used, children passed the task only at the age of 5. It follows that there are children – exactly seven – who passed at least one of the FBTs but did not pass the Memory for complements task. This data in itself already challenges de Villiers' hypothesis, since children passed the memory for complements later than the FBTs, thus it cannot be the prerequisite of false belief understanding. These data are in line with data from other languages, in German or in Mandarin Chinese and queries whether the relation found by de Villiers between FBT and complements is universal across languages.

Further investigating the relation between the three language tests and the FBTs, correlations were found between the verbal FBT and 3 of the language tests; Peabody, TROG-H, Memory

for complements. More interestingly, no correlation was found between any of the language tests and the nonverbal FBT. This, again is in line with the findings of Study 1, that relations found between language tests and FBT – both in this study and in the literature – are due to the verbality of the FBT. This is certainly a further challenge of the Complement hypothesis and in more general of the hypotheses that suggest that it is language, or a certain aspect of it that is necessary to pass FBT.

However the relations with the verbal FBT suggest that at least if false belief understanding is measured with a verbal test it is possible that a certain level of language is required from children to pass the verbal FBT. We found that after controlling for the Peabody the correlation with the TROG-H was still present unlike the correlation with the memory for complements. This suggests that syntax – but not a specific aspect of it – has a special relation with FB understanding. This result is somewhat in line with Astington & Jenkins findings (1999), who interpreted their results that syntax has a unique role in FB understanding. However, since my data is not longitudinal as theirs was, and I have only correlational data, the direction of the relation can be either way, or even a third ability can be responsible of the correlation.

Children with ASD

Children with ASD have deficit in different aspects of ToM including FB understanding. Still some high(er) functioning children with ASD still pass FB tests, however a close relation was found between these children' s verbal MA and their performance on FBT. A relatively high level of language ability (above the verbal MA of 7) is needed for children with ASD to pass FBT. Thus, an even stronger relation was expected between their performance on language tests and FBT performance. Helen Tager-Flusberg based on these data hypothesized and found relation between a more advanced form of complement task (CIW task) and verbal FBT also Lind & Bowler found the same relation although on a very heterogeneous population.

The results of *Study 3* are perfectly in line with these data. 'Relatively higher functioning' children with ASD showed a strong and significant relation between their performance on the verbal FBT and all three language tests, the association was stronger than the one found in typically developing children. Moreover, even after controlling for the effect of vocabulary,

the relation with the CIW was still strong and significant. This finding is in line with Happe's assumption; children with ASD might use verbal compensation strategy when they pass verbal FBT.

Also unlike in Senju and colleagues (in press) study, since children's behavior was measured with acting out behavior, it is not likely these children's poor performance on the nonverbal FBT was due to the lack of motivation. It is more likely that is represents their deficit in false belief understanding.

5.4. Study 4: Developmental relationships of understanding complements, naïve theory of mind and word acquisition

5.4.1. Background

In a recent study, Happé & Loth (2002) tested the Sperber and Wilson's (Sperber, 2000; Sperber & Wilson 2002) hypothesis of distinct sub-module within theory of mind or as they refer to the ability, mind-reading. Sperber and Wilson (2002) outline two possible ways of intentional communication, the first that is an "application of a general mind-reading module to the problem of identifying the speaker's meaning (a neo-Gricean view). Second, it might involve a sub-module of the mind-reading module, an automatic application of a relevance-based procedure to ostensive stimuli, and in particular to linguistic utterances " (p. 16). As we saw it in part 3.2.2 they argue with the second assumption.

Happé and Loth (2002) tested the developmental aspect of Sperber and Wilson's theory by creating a word-learning FBT; children had to track a false belief in order to learn a novel word (see 5.4.4. for the description of the task). They found that in spite of the increased task complexity, significantly more 3-5 year-old children passed the false belief task when it was combined with a word-learning task, than when presented in its standard form. Happé and Loth – after excluding other possible explanations of their findings – interpreted these results that they support Sperber and Wilson's hypothesis; the theory of mind mechanism might be not a unitary mechanism but it might consist of more – at least two – component mechanisms; one for inferring the communicator's intentions and the other is for inferring the actor's intentions, moreover developmental trajectories of these components may be different.

These findings are not only a new assumption of theory of mind module, but also they challenge the de Villiers's complement hypothesis. If there is version of the standard FBT (word-learning FBT) based on metarepresentation that children can pass earlier, then it might be another argument to reject the complement hypothesis. However, even if it is found, that children pass the word-learning FBT earlier than the complement task (Memory for Complements), one could argue that similar modification of the complement task, namely embedding it into a word-learning context, could reduce the age of the passers, just like in the case of the word-learning FBT[20]. Thus I tried to make the two tasks; word-learning FBT and

[20] Note that such a criticism would probably come from an alternative interpretation of the Happé & Loth (2002) study; the word-learning FBT has less cognitive (e.g. executive function) demand, than the standard FBT.

Memory for Complements task make as similar in this respect as possible by embedding them into a word-leaning context. That was the motivation of creating the word-learning Complement task (WLCT). However, if it is found that by embedding the word-learning context into the complement task, more children pass it than the standard Memory for Complements task, than it questions Happé & Loth's interpretation of their results, and suggest that the word-learning context makes – at least these two – tasks easier (either by making the task more interesting or by decreasing the executive demands).

5.4.2. Hypotheses and Questions

(1) The same results are expected to found with word-learning FBT as in Happé & Loth's study; significantly more 3-5-year-old will pass it than the standard FBT.

However, the main question of the study was the following:
(2) Can children pass the word-learning false belief task earlier than the complement task?
(3) Can children pass the word-learning complement task earlier than the standard complement task?
(4) The *extended de Villiers hypothesis:* the mastery of sentential complements predicts verbal false belief understanding not only in the standard, but also in the word-learning context.
(5) The better performance in false belief understanding in word learning contexts is due to a more general effect of this context, which is either decreasing the cognitive demands of the task or to increase the attention of the children - e.g. by making the task more interesting – and not to the different developmental trajectories of two separate mechanisms of theory of mind.

5.4.3. Participants

Sixty-five *typically developing children* aged 2.5 to 5.5 years were recruited from 4 different local preschools in Budapest, Hungary. Fourteen children were excluded as they failed the memory pretest: twelve 3-year-olds, one 4-year-old and one 5-year-old (see description at Materials). Thus 51 children (26 girls) were included in the final sample (and were divided into four groups: 2.5-3 years: 5 children (the low number of the youngest children is due to

the floor effect was found at this age); 3-4 years: 15 children; 4-5 years: 16 children; 5-5.5 years: 15 children) their mean age was 4;2 years.

5.4.4. Materials and procedure

Each child was tested individually in a quiet room or a separated area of the nursery school. Children were seated opposite the experimenter at a table.

Memory pretest

A toy tiger was shown to the child and was asked: "What is that?" After the child had named the toy, the tiger was put into a box and closed it. Then a toy lion was shown to the child and was asked: "What is that?" Then the experimenter took the tiger out of the box and put the lion inside. The child was then asked: "What is in the box now?" and "What was in the box in the beginning?" Only those children who passed both questions proceeded to the main investigation. In contrast to the original pretest used by Happé and Loth the two toys in this study's pretest were visually very similar to minimize the possibility that in later tests children would fail because they mixed the objects due to the visual similarity. Probably that is the reason why we had to exclude much more children from the investigation (14 children) than Happé and Loth (only 5 children) did.

Language tests

Two language tasks were the same as in Study 3: *Peabody Picture Vocabulary Test* (PPVT; Csányi, 1974), *Trog* (Bishop, 1983).

Memory for Complements task: it was the same as in Study3, except that since children were tested from the age of 2.5, there were fewer tasks in it, namely 4 instead of 12, 3 false and 1 true complement.

Word-learning complement task (Figure 19): the standard Memory for Complements task was embedded in a word-learning context. An example: 'She said to the girl that there was a bottle in her hand but it really was a TIMA. What did she say?'

The correct answer was 'that there was a bottle in her hand' ('bottle' was also accepted) (In Hungarian: "Ő azt mondta a lánynak, hogy egy üveg van a kezében, de igazából egy TIMA volt benne. Mit mondott a lánynak?"). The sentences were accompanied with pictures, and the

experimenter always pointed to the relevant picture, just like in the Memory for Complement task. There were two new objects in each second picture, so if the child did not follow the story with the pictures she/he was unable to decide which of the two new objects was labeled. Children did not have to repeat the novel word, which would have meant extra difficulty for them. (Note, that the children do not have to name the new object in the word-learning FBT either, see next paragraph). But just like in the word-learning FBT it was tested if the child can decide to which of the two novel objects referred the new label, following Baldwin's (1993) procedure. (Find the entire test with pictures in the Appendix 4). The child was asked the following test questions:

Q1: 'Show me the tima!?'

Q2: 'Which one is the tima: Is this one here the tima (pointing at A/B randomised)?'

Q3: 'Is this one the tima (pointing at A/B randomized)?'

Q4: 'Can you give me the tima?'

In addition 3 control questions were also asked, 2 memory questions: 'Which one was really in her hand?' and 'Which one was on the ground?' and 1 preference question: 'Which one is the best? or Which is your favorite?' No reference was found to any of the objects; 34% of the children preferred the labeled toy, 37% preferred the unlabeled toy and 29% liked both objects equally.

Children were presented four word-learning complement tasks; 3 false and 1 true.

Unlike the word-learning FBT, the word-learning Complement task tests word-learning and the mastery of complements in two separate tests. Although the more questions (there is one more question in this task compared to word-learning FBT which tests the acquisition of the complement) clearly decreases the chance of passing the test, the separate testing can give us an inside of the different background mechanisms. Thus three scores will be reported of this task, the complement score, the word learning score and a combined score (whether or not passed both scores). Children passed the complement part if they passed at least 3 of the 4 trials, including the true complement trial. Children passed the word-learning part if they answered at least 3 of the 4 questions correctly, and only if the one question they failed was the prompted question 'Is this one the modi?'.

- [Pointing to Picture 1] She said to the girl that there was a bottle in her hand, [Pointing to Picture 2] but it really was a TIMA
- Q: What did she say?

Figure 19. The Word-learning Complement Task

Theory of Mind

One *verbal FBT* was administered the same way as written in the earlier studies; the one verbal FBT was *the Sally Anne test*.

Word learning FBT (Happé & Loth, 2002) (Figure 20a & 20b): For the word-learning tasks, two pairs of novel, un-nameable objects were used, chosen to be attractive for children and visually distinct from one another. Assignment of puppets and object pairs to the false and true belief conditions was counterbalanced, as were non-words used ('modi', 'wug').

Sally places a new object in a box then she leaves. Anne comes in and she has another new object and puts her object into the box (takes Sally's object out of it). Anne leaves too. Sally comes back and labels the object three times without opening the box 'Do you want to see the wug? There is a wug in the box! Let's see the wug!' Then the experimenter displayed both objects in front of the child tests if the child has attached the new label to the correct reference. This was done by using Baldwin's (1993) method, just like above, the same way as used in Happé and Loth's study. In addition 3 control questions were also asked from the children; 2 memory questions: 'Which one is Sally's and which one is Anne's? Which one was in the box in the beginning?' and 1 preference question: 'Which one is the best? or Which is your favorite?' to make sure that children did not pick one of the two new objects because they liked it better. No effect of preference was found; in word-learning FBT 39% of the children preferred the labeled object, 36% preferred the unlabeled object, and 25% liked both objects equally. In the word-learning TBT 37% of the children preferred the labeled object, 38% preferred the unlabeled object and 25% liked both toys equally.

Children were also tested with the true belief version of the word-learning test, in which Anne puts her new object into the box in the presence of Sally so the new label refers to Anne's object, which in this case really is in the box. The reason of this to avoid that e.g. children pass the task based on simple association between the first puppet and her new object. However, during both word-learning belief tasks, right after the switch of the objects the child was asked if the first puppet, Sally saw what Anne just did. If the child's answer was incorrect, he/she was explicitly corrected, e.g. 'No, she did not see it because she was not here'. Children passed the task if they answered at least 3 of the 4 questions correctly, and only if the one question they failed was the prompted question 'Is this one the modi?'. (The procedure of the word-learning FBT was the same as the one used by Happé and Loth, 2002.)

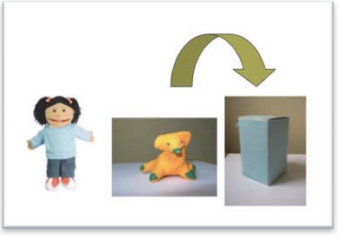


Figure 20a. The word-learning FBT; Sally places a new object into the box.

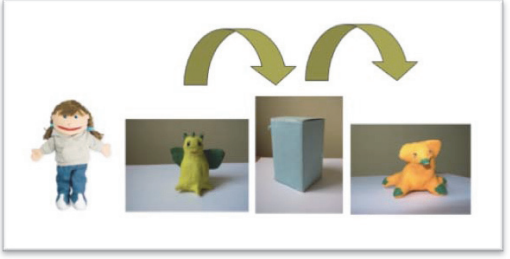


Figure 20b. The word-learning FBT; Anne replaces Sally's new object into her new object in the box.

5.4.5. Results

Thirty-eight children passed all control questions in *Sally Anne FBT*. Of these, 27 children, so the 71% passed the test question of the Sally Anne FBT too. In the word-learning FBT thirty children passed all the control questions and only 10 children (33%) answered the test

questions correctly. Finally in the word-learning true belief test (TBT) thirty-eight children passed the control questions and 22 of them (57%) also passed the test questions.
To compare the performance on the FBT to the performance on the word-learning FBT, I selected the children who passed the control questions for both tasks (Table 10), this left 30 children in the analysis. Using McNemar test it was found that significantly more children passed the Sally-Anne FBT than the word-learning FBT (chi square=15.53, df 1, $p \leq 0.001$). Following Happé and Loth's statistical analysis I also focused on the children who could pass only one of these two tests. Of the 10 children who passed the word-learning FBT only 1 (10%) failed the Sally-Anne task. In contrast, of the 22 children who passed the Sally-Anne test 13 children (59%) failed the word-learning FBT.
However, just like in Happé & Loth's study, there was no difference between the word-learning TBT (38/22) and the word-learning FBT (30/10) (chi square=0.89, df 1, $p > 0.05$)

		Word-learning FBT		
		Pass	Fail	
FBT (Sally-Anne)	Pass	9	13	∑: **22**
	Fail	1	7	∑:8
		∑: **10**	∑:20	

Table 10. Contingency table showing numbers of children passing and failing the Word-learning False Belief test (FBT) and the standard Sally-Anne false belief test. (Excluding the children who did not pass the control questions for both tasks, n=30.)

Since our results are just the opposite as Happé and Loth's results further analyses were concentrating on the word-learning FBT to determine at what age children can pass it in contrast to the Sally Anne FBT. Children aged 2.5 to 3 years had a floor effect in every test so their results are not shown in the diagram.

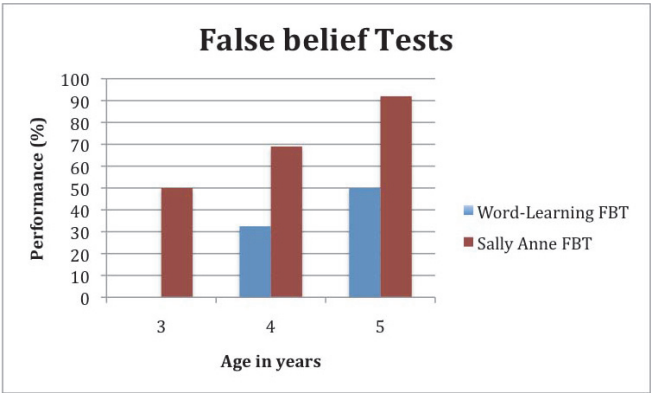


Figure 21. Children's performance on the word-learning FBT and the Sally Anne FBT as a function of age.

Figure 21 presents that only 50% (12/6) of the oldest children – who aged 5 to 5.5 years – passed the word-learning FBT compared to the 92% who passed the Sally Anne FBT. It means it was much more difficult for them than the Sally Anne FBT. The results on the Sally Anne FBT fit into the results of the literature that children pass the false belief tests from about the age of 4.

Fifty-one children passed all control questions of the *Complement task*. Of these, 26 children, so 51% passed the test questions, too. In the *word-learning complement task* forty-four children passed all the control questions, 22 (50%) of them passed the complement test question, 26 (60%)the word-learning test questions and 14 (32%) passed both tasks and therefore the word-learning complement task as a whole too.

To compare the performance on Complement task to the performance on word-learning complement task I selected the children who passed the control questions for both tasks, which left 43 children in the analysis (Table 11). Using chi square tests it was found that significantly more children passed the Complement task than the word-learning complement task (chi square=10.28, df 1, $p \leq 0.005$). I used the same method of comparison as in the FBT thus I focused on the children who could pass only one of these two tests. Of the 14 children who passed the word-learning complement task only 1 (10%) failed the standard Complement task. In contrast, of the 25 children who passed the Complement test 12 children (48%) failed the word-learning FBT.

		Word-learning Complement task		
		Pass	Fail	
Complement task	Pass	13	12	∑: **25**
	Fail	1	17	∑:18
		∑: **14**	∑:29	

Table 11. Contingency table showing numbers of children passing and failing the Word-learning Complement Task and the standard Complement task (excluding the children who didn't pass the control questions for both tasks, n=44).

However, it worth taking a closer look at the other scores of the *word-learning complement task*. As I said 50% of the children passed the complement question in the task, which is very similar to the performance on the standard complement test. Note that children can pass this part without knowing anything about the novel word.

The picture gets even more interesting if we check how children performed in the 3 different age ranges. Seven (46%) 3-year-olds passed the word-learning part of the task and only 1 (6.7%) passed the complement part. Nine (56%) 4-year-olds passed the word-learning part and 7 (44%) passed the complement part, among these children only 4 passed the word-learning part too. And finally, ten (67%) 5-year-olds passed the word-learning part and 14 (93%) passed the complement part.
It looks like, when children get to the age when they are able to pass the complement test, they focus on that task and perform at a very high level (93%). Certainly there is another, maybe not unrelated explanation too, that it is due to the order of the test questions; it was always the complement question that was asked first, so answering the questions of the word-learning part needed bigger working memory load.

After comparing the two complement tests, the standard and the word-learning one, it becomes clear that the significant difference found between them is due to the 3-and 4-year-olds poor performance on the word-learning complement task. Unlike in the word-learning FBT, children's performance on word-learning complement task reached a high level by the

age of 5. Moreover, the performance on the word-learning complement task is similar to the standard complement task, certainly no significant difference was found at this age between them (McNemar, $p>0.05$).

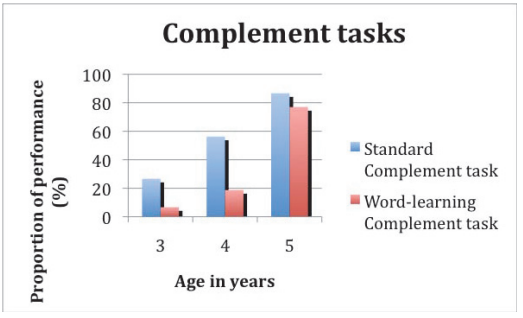


Figure 22. Children's performance on the word-learning complement task and the standard complement task as a function of age.

Children performed on the standard complement task similarly as in Study 2, although a shorter version of the task was used. At the age of 3 children performed under chance, but at the age of 4 they performed above chance, around 60%, and at the age of 5 their performance was close to ceiling.

Since further hypotheses were built upon the results of Happé and Loth, that I could not reproduce – moreover, I found opposite results – there was no reason to test the relation between the word-learning FBT and the complement tasks.

5.4.6. Conclusion

I would like to stress that what was found was not 'simply' a lack of result, but a robust opposite result of what Happé and Loth found and not only in one test but in two. In other words what I found was not only that word-learning FBT is not easier than the original FBT, but that when children already pass the original FBT, they still do not pass word-learning FBT above chance. We also know that children as early as 18 months are capable of tracking the intended referent of a novel word, even if that referent is inside an opaque container at the time of labeling (Baldwin, 1993). Thus children around the age 4 are able to pass the 2 tests separately but could not pass the test, which combined these two tests. There are two related

questions that have to be considered. The first is why children found the tests, which were embedded into a word-learning context so difficult. And the second is why our results were just the opposite as Happé & Loth's results.

There are a couple of possible answers to the first question. One of the possibilities is that the word-learning context made both the FBT and the Complement task even more complex. It is easy to argue that the combination of two tasks – word learning and FBT or Complement – means additional cognitive load. The other possibility is that the different task difficulty is due to the different chance rates to pass the tests. In the standard FBT children has about 50%[21] chance to pass the test (1 test question, 2 possible answers), while in the word-learning FBT kids have to answer 3 out of 4 questions correctly, in order the pass the test. That means only 3 of the 16 possible patterns of answer would result in a pass, which is 18.75% probability. Similar difference between the chance rates can be found between the standard Complement task (around 50%) and the word-learning complement task (16%). Thus children have more chance to pass the standard FBT and Complement task than their word-learning versions. However, since the tests were the same, the chance rates of passing the FBTs were the same in the two studies, just like the cognitive capacity needed to pass the tests. Still, the results of the two studies are just the opposite. The second question is what could lead to this difference?

A surprising data of the Happé & Loth study is that although in most studies children pass the standard FBT around the age of 4, in this study only 36% of the 3-5-year-olds passed it. The reason of low level of performance is unknown, though the authors speculate that maybe the standard FBT was not interesting for the children (my speculation is that maybe they were presented with the exact same standard FBT task recently by other research groups – this happens in kindergartens close to universities). But this data only explains the half of the results, but still leaves unanswered the question why and how 87% of the children passed the word-learning FBT in the Happé & Loth study in contrast to the 33% found in my study. The only explanation I can think of is the children tested in their study represented a 'special sample' and were very much tuned to word-learning tasks due to either a special education in the kindergarten or to special family background, maybe they even participated in a similar word-learning context shortly before they were tested. But this is certainly, just speculation. However, the data that the sample in my study performed on the standard FBT in line with the literature data suggests that my sample was rather an average, typical sample than the one in

[21] Not exactly 50% since children are biased to the location where the object below the age of around 4 when the consistently start to pass the test.

the Happé & Loth study. There is certainly one more difference between the two studies, which is the language these studies were administered, and therefore the mother tongue of the children participated. Nevertheless, I am not aware of any reason why the word-learning FBT, but not the standard FBT would be more difficult in Hungarian than in English. The linguistic complexity of the two FBTs is rather similar, so in case Hungarian language made the word-learning FBT more difficult, I should have found a similar effect on the standard FBT. In sum, I do not see enough argument to believe that the difference between the two word-learning FBTs – in my study and in Happé & Loth study – was caused by the different languages they were administered.

Interestingly, however a phenomenon mentioned in Study2, returned in this study. Happé & Loth found that the word-learning TBT (true belief test) was slightly more difficult than the word-learning FBT[22]. Although the difference was not significant, it clearly resonates with the findings of Study2, although in this test the experimenter makes sure that the child knows what the first character can or can not see and ask him/her while the second character switches the toy, if the first character can see what the second character is doing. The experimenter corrects the child if necessary. Thus it is very improbable that children who failed this test got confused what the first character saw. The authors explains their results as follows:

"Whilst young children often delight in stories including trickery (e.g. Sullivan and Winner, 1993), the true belief version essentially consisted of two characters presenting their toys one after the other. Hence, as the to-ing and fro-ing of the first character was less significant for the subsequent comprehension of the story plot, her actions may have appeared less motivated and plausible, resulting in confusion or inattention for some children " (p.30).

As we can see their argument is very similar to mine in Study2. I would like to add that according to my data, this phenomenon is more likely at the age when children just become able to understand false beliefs, and not at an earlier age. In my sample, although children found the word-learning TBT somewhat easier, than its FBT version, the difference was not significant, just like in original paper. Also note, that children in my study did not pass the word-learning FBT above chance, thus following the authors' arguments, this could be the reason of their somewhat better performance on the word-learning TBT, but the difference, again, is not significant.

[22] There was no standard TBT administered in their study.

In summary, in opposition to Happé & Loth's result, significantly more children passed the standard FBT than the word-learning FBT, thus the word-learning FBT is more difficult than the standard FBT. Moreover, children who passed only one of the two tasks were more likely to pass the standard FBT and fail the word-learning FBT than vice versa. Since there was no remarkable difference between the methods of the two studies it is suggested that it is due to different sample of children tested in the studies, more specifically that the children in Happé & Loth's study were was worse in the standard FBT and better in the word-learning FBT than the average typical children. In light of the new data, there is no reason to assume different ToMM to explain these data.

Chapter 6. General conclusions

The aim of the Dissertation was to study the developmental connection between language and FBT in the kindergarten age by exploring the relation between different aspects of language and FB understanding in both typical and atypical populations.

A crucial element of this exploration was to develop a nonverbal FBT. A constant source of criticism of studies exploring this developmental relation is due to the fact that the FB understanding is accessed by a verbal test and the influence of its verbality cannot be excluded. In other words, the relation found in these studies might be at least in part due to the verbality of the FBT. *Study 1* presented the validation of the new nonverbal FBT, which as I discussed in Chapter 2 and *Study 1* has advantages when compared to other already existing FBTs. One of the most important findings of the validation process was the similar developmental trajectory found between the verbal and nonverbal FBT. Children from the age of 4-5 are able to pass both verbal and nonverbal FBT. This result is in accordance with literature data. Another important finding was that there was no correlation between the two FBTs when the effect of age was controlled for. However, when only the location change FB trials were left in the analysis, there was a tendency between the two tests. In *Study 1* I discussed a couple of possible explanations – including the one that the new nonverbal FBT is not a valid method of accessing false belief understanding –, and I argued that the lack of correlation could be due to the inconsistency of children's performance on the test. When only those children were entered into the analysis that either passed or failed all of the trials within the FBTs, the relation between the two tests became statistically significant. The arguments that nonverbal FBT is a valid method of measuring false belief understanding was further strengthened in *Study 3*, where, although with a smaller sample size, the correlation was significant between the verbal and nonverbal FBTs.

Study 2 focused on a possible criticism of *Study 1*, namely, that in theory the nonverbal FBT could be passed by a simple strategy based on an association between the characters and their hiding places. Although this is a valid criticism of a remarkable proportion of the studies of verbal FBT, a location change TB trial was added to both the verbal and the nonverbal FBT. While children's performance on the FB trials and the non-location-change TB trials did not decrease (actually by shortening the nonverbal FBT it increased), the performance on the

location change trial was below chance. The difference between the location change TB trial and the FB trials was not significant, but reached significance between the location change and no location change TB trials. While these results are somewhat surprising, they are not without precedent. Lohmann et al. found very similar results that we found at the 3-year-olds. Another example in the literature is Happé & Loth's study. Although they used a modified version of the true and false belief tests (the word-learning belief tests) they found the same results at 3-5 year-old children, namely, the FBT was easer for the kids than the TBT. Unfortunately Happé & Loth did not report more precise data about how the 3-,4-, and 5-year-olds performed independently, therefore we cannot follow the developmental trajectory of these children's performance. Roth & Leslie (1998) found the same phenomena using the TBT that younger children performed better on the task than older children around the age of 4., however the difference did not reach significance neither in our study nor in theirs. In summary, we argue that the new nonverbal FBT is a valid measurement of FB understanding, and therefore the version presented in *Study 1* will be used in the further studies.

A really interesting result was obtained for *Study 1*, which is the focus of the Dissertation: the relations found between the language tests and the FBTs. While we found a positive correlative relation between the language tests and the verbal FBT, just like the majority of the studies do, there was no relation between the language tests and the nonverbal FBT.

The aim of testing children with developmental language impairments was twofold; (i) to test their false belief understanding, since the research of this field is very limited, and the few studies were done mostly concentrate on SLI population and, (ii) to explore the developmental relation between language and FB understanding in a population in which language is impaired. Our data confirmed the delay in FB understanding that was found in the research. Five to 7;7-year-old Children with DLI did not pass either the verbal or the non-verbal FBT significantly above chance. Also, based on earlier research (Miller, 2001) it was expected that children with DLI would perform better if the linguistic demand of the task is reduced, that is, they would perform better on the non-verbal than on the verbal FBT. In contrast, it was found that children with DLI performed somewhat better on the verbal FBT, although the difference was not significant. After creating more homogeneous subgroups with statistical method within the sample, a more and a less advanced subgroups were found with significant differences between both of FBTs and language tests. Note that the IQ and (nonverbal) mental age was not different, thus the other differences were not simply due to general intellectual differences. Interestingly, no relation was found within the 2 subgroups between false belief understanding (verbal and nonverbal) and language ability, moreover

between the two language tests either. This suggests an atypical pattern of development within which these capacities develop independently.

The results of *Study 1* on typically developing children were replicated in *Study 2*, while the focus of the study was to test de Villiers' hypothesis if sentential complements have a uniquely predict later FB understanding. We did find a correlative relation between all of the language tests, including the complement test and verbal FBT, however there were a couple of children – more than in de Villiers study – who did pass the verbal FBT but performed poorly on the complement task. This is not surprising if we consider that Hungarian children in our study passed the complement task later as it was indicated by de Villiers earlier studies with English speaking children. Moreover, when the effect of the Peabody test, which measures children's vocabulary and strongly correlates with general language ability was controlled for, the only correlation that was still found was with children's grammatical ability. These results agree with Astington and Jenkins (1999) findings, however they also warn that these results might be due to the verbality of the FBT. Indeed, when the FB understanding was tested with the nonverbal FBT, we did not find any relations between the FBT and the language tests, neither with the TROG-H, nor with the complement task.

Children with ASD were also tested with this paradigm, since they represent a very relevant population in this question. Children with ASD have difficulties in passing FBT. It has been argued that there is a strong relation between their language ability and their performance on verbal FBT, namely, that they do not pass such tests under the verbal mental age seven (Happé, 1995, Fisher, 2005). Recent research found that the poor performance under this verbal mental age is not due to the verbality test, because even when FBT was presented completely nonverbally, ASD children still performed at chance (Colle et al, 2007; Senju et al, in press). Our results are in line with these findings; older 'higher functioning' children with ASD performed at chance on the nonverbal FBT. Nevertheless, their performance on the verbal FBT was better, however not significantly better. Moreover, we found a very strong relationship between the language tests, especially between the complement task (CIW) and the verbal FBT, but only with the verbal FBT and not with the nonverbal FBT, the implication of which can be twofold: (i) it could suggest that children with ASD who pass the verbal FBT used verbally mediated compensatory strategy (ii) de Villiers complement hypothesis is not tenable for the following reasons. This latter is only true if ToM capacity is tested with a verbal FBT, suggesting that it is only a byproduct of the verbality of the test, and only in the ASD population, in which, when compared to typically developing children, a stronger relation was found between their language ability and ToM capacity (Happé , 1995; Fisher et

al, 2005). Also note that the relation found in this study was only correlational, which does not indicate a causal relation, however this possibility cannot be rejected. In summary, the de Villliers hypothesis is not tenable either in typically developing children or in the ASD children. In both samples any relations between the mastery of complements and the performance on the verbal FBT is due to the verbality of the test – since no such a relation was found with the nonverbal FBT –, and does not reflect an essential causal relation. However, somewhat different patterns were found within the two samples regarding of the key aspect of language responsible for this byproduct. In typically developing children it is general grammar comprehension, while in children with ASD it is the mastery of complements. Unfortunately *Study 4* could not add to this picture on the merits. Since Happé and Loth's results were not replicable with the word-learning FBT, our further hypotheses with the word-learning complement task were not testable.

However it is important to note that at least a few kids with ASD (7/16) did pass the nonverbal FBT. They either used some alternative, non-mentalistic strategy or used the ToM. As we saw it in *Study 1* and *Study 2* such a possible strategy could be based on the association between the character (puppet) and its hiding place. At this point this possibility cannot be excluded, however considering these children's age and intellectual level (mean IQ around 80) and also that only a relatively small number of children could pass the test, this alternative strategy does not seem to be trivial for children. But even if they used ToM, it was a structured test situation, which does not reflect the complexity of the real life's social situations. Therefore it is a further question whether the performance on the nonverbal FBT relates to these ASD kids social skills in the everyday-life. However, the correlation found between the nonverbal FBT and the ADOS (Stefanik, 2005; Győri et al, 2007) suggests that there might be such a relation, but it needs to be further explored.

And finally, as mentioned above the better performance on the word-learning FBT compared to the standard FBT could not be replicated on children at kindergarten age. If we accept these data over Happé and Loth's results due to the arguments I made in the conclusion part of *Study 4*, we have no reason to question the unitarity of ToM capacity, at least not because of these findings – note that our facilitation hypothesis was not supported either. The opposite results, the word-learning FBT is more difficult than the standard FBT can be explained by a more parsimonious way; the higher complexity of the word-learning FBT led to this difference.

The aim of the study was to get closer to the question what the developmental relation between false belief understanding and language is during the kindergarten age. I would like

to emphasize that if we consider this developmental relation in a broader age range, the relation seem to be bidirectional. For instance, it has been argued that joint attention behavior is critical for early word learning (e.g. Baldwin, 1993). Recent interesting research found false belief understanding, at least on some level, around or even under the age of 2. However, further research is needed to explore how the new FBTs using gaze responses can measure false belief understanding during later development (note that there is some data from Senju et al (in press) that argues that at least the one test developed by Southgate et al is a sensitive test even around the age of 7.). It is also the question of future research to explore the kind of relation that would be found between language and FB understanding accessed with these tests. In the Dissertation we could see how a nonverbal FBT, that requires acting out, modifies the earlier findings regarding the relation in question. If we accept our new nonverbal FBT as a valid test to measure FB understanding, we can conclude that the relation(s) some of the earlier studies found was due to the verbality of the FBT and does not reflect an essential causal relation. As we saw it in Chapter 2 however the approach of 'no special role for language in Tom development' have different perspectives too. Our results could be clearly explained by a nativist modular approach; both language and ToM are innately specified, thus the results suggest that by eliminitating/canceling the verbality of the ToM the performance limitations decreased. That is, the better performance found with the nonverbal FBT still supports this approach, however these findings were not consistent. (CONSISTENT??).
However, as we saw our findings do not support all kinds of modularity thesis. Sperber and Wilson (2002) proposed a unique role of language; developments that are requisite for effective conversation are encapsulated within the language module; thus the child might be able to use ToM knowledge appropriately in language tasks, but not in other tasks outside of language. Our findings are clearly in opposition with this approach.
Also note, however, one does not necessarily need to assume either innateness or modularity to explain our data. For instance, Chandler et al (1989) claim that the role of language is only superficial since passing a standard FBT requires a certain level of language ability, but they do not assume an adult-like competence in the infant's mind.
Finally, since the Dissertation focuses strictly on the two abilities in question, the possibility of another, third factor's essential role in ToM development still cannot be excluded. These are interesting questions and ideas that further research needs to explore.

Chapter 7. References

APA (1994). *Diagnostic and Statistical Manual of Mental Disorders* (*4th ed)*. Washington (DC): American Psychiatric Association.

Astington, J. W. (1994). *The child's discovery of the mind.* Fontana Press, London.

Astington, J. W. (2001). The future of theory-of-mind research: Understanding motivational states, the role of language, and real-world consequences. *Child Development, 72*(3), 685-687.

Astington, J. W. & Baird, J.A. (2005). Introduction: Why language matters. In: Astington, J.W. & Baird, J.A. (Eds). *Why language matters for theory of mind.* (pp3-25). New York: Oxford University Press.

Astington, J.W. & Baird, J.A. (2005). Representational development and false-belief understanding. In: Astington, J.W. & Baird, J.A. (Eds). *Why language matters for theory of mind.* (pp3-25). New York: Oxford University Press.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Dev Psychol, 35*(5), 1311-1320.

Baldwin, D.A. (1993). Early referential understanding: infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29,* 832–843.

Baron-Cohen, S. (1989). The theory of mind hypothesis of autism: a reply to Boucher. *Br J Disord Commun, 24*(2), 199-200.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37-46.

Baron-Cohen S, Tager-Flusberg H, Cohen DJ, (Eds). 2000. *Understanding Other Minds*. London/New York: Oxford Univ. Press

Bartsch, K., & Wellman, H. (1989). Young Childrens Attribution of Action to Beliefs and Desires. *Child Development, 60*(4), 946-964.

Bishop, D.V.M. (1983). *The Test for Reception of Grammar*. Age and Cognitive Performance Research Centre, University of Manchester.

Bishop, D. (1997). *Uncommon understanding: Development and disorders of comprehension in children.* Hove,UK, Psychology Press.

Bloom, P. (1999). Theories of word learning: Rationalist alternatives to associationism. In

W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of Child Language Acquisition* (pp. 249-28). San Diego, CA: Academic Press.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*(1), B25-31.

Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when and why. *Cognitive Psychology, 6,* 380–420.

Boucher, J. (2003). Language development in autism. *Int J Pediatr Otorhinolaryngol, 67 Suppl 1*, S159-163.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development, 70*(2), 381-395.

Carpendale, J. & Lewis, C. (2006). *How children develop social understanding*. (Eds.), Oxford: Blackwell Publishing.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr Soc Res Child Dev, 63*(4), i-vi, 1-143.

Chandler, M. J., & Carpendale, J. I. M. (1998). Inching toward a mature theory of mind. In M. Ferrari & R. J. Sternberg (Eds.), *Self-awareness: Its nature and development (*pp. 148–190). New York: Guilford Press.

Chandler, M.J. Fritz, A.S., & Hala, S.M. (1989). Small scale deceit: Deception as a marker of 2-, 3- and 4-year-olds' early theories of mind. *Child Development, 60,* 1263-1277.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A. & Drew, A. (2000). Testing joint attention, imitation and play as infancy precursors to language and theory of mind. *Cognitive Development, 15*, 481-498.

Cheung, H., Hsuan-Chih, C., Creed, N., Ng, L., Wang, S. P., & Mo, L. (2004). Relative roles of general and complementation language in theory-of-mind development: Evidence from Cantonese and English. *Child Development, 75*(4), 1155-1170.

Clements, W. A., & Perner, J. (1994). Implicit Understanding of Belief. *Cognitive Development, 9*(4), 377-395.

Cohen, L. B. (2004). Uses and misuses of habituation and related preference paradigms. *Infant and Child Development, 13*, 349-352.

Colle, L., Baron-Cohen, S., & Hill, J. (2007). Do children with autism have a theory of mind? A non-verbal test of autism vs. specific language impairment. *J Autism Dev Disord, 37*(4), 716-723.

Colonnesi, C., Rieffe, C., Koops, W., & Perucchini, P. (2008). Precursors of a theory of mind: A longitudinal study. *British Journal of Developmental Psychology, 26*(4), 561-577.

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology, 12*(3), 349-358.

Conti-Ramsden, G., Crutchley, A., & Botting, N. (1997). The extent to which psychometric tests differentiate subgroups of children with SLI. *J Speech Lang Hear Res, 40*(4), 765-777.

Csányi, F.I. (1974). *Peabody Szókincs Teszt*. Bárczi Gusztáv Gyógypedagógiai Főiskola. Budapest.

Csépe, V. (2005). *Kognitív fejlődésneuropszichológia*. Budapest: Gondolat Kiadó

Davis, H. L., & Pratt, C. (1995). The Development of Childrens Theory of Mind - the Working-Memory Explanation. *Australian Journal of Psychology, 47*(1), 25-31.

Dawson, G., Osterling, J., Meltzoff, A. N., & Kuhl, P. (2000). Case study of the development of an infant with autism from birth to two years of age. *Journal of Applied Developmental Psychology, 21*(3), 299-313.

de Villiers, J.G. (2005). Can language acquisition give a point of view? In: Astington, J.W. & Baird, J.A. (Eds). *Why language matters for theory of mind.* (pp3-25). New York: Oxford University Press.

de Villiers, J.G. (2007) The interface of language and theory of mind. *Lingua, 117,* 1858-1878.

de Villiers, J. & Pyers, J. (1997). Complementing cognition: The relationship between language and theory of mind. Proceedings of the 21st Annual Boston University Conference on Language Development. Somerville, MA: Cascadilla Press.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*(1), 1037-1060.

de Villiers, J., Roeper, T., & Vainikka, A. (1990). The acquisition of longdistance rules. In: L. Frazier & J. de Villiers (Eds.), *Language processing and acquisition*. Dordrecht, Netherlands: Kluwer.

Dennett, D. (1978). Response to Premack and Woodruff: Does the chimpanzee have a theory of mind. *Behavioral and Brain Sciences, 4*, 568.

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language, 106*(2), 144-152.

Dunn, J. & Brophy, M. (2005). Communication, relationships, and individual differences in children's understanding of mind. In: Astington, J.W. & Baird, J.A. (Eds). *Why language matters for theory of mind.* (pp3-25). New York: Oxford University Press.

Dunn, J., Brown, J., & Beardsall, L. (1991). Family Talk About Feeling States and Childrens Later Understanding of Others Emotions. *Developmental Psychology, 27*(3), 448-455.

Dunn, L. M., & Dunn, L. M. (1997). *Examiner's Manual for the PPVT*-III: Peabody Picture Vocabulary Test Third Edition. Circle Pines, MN: American Guidance Service.

Farrant, B. M., Fletcher, J., & Maybery, M. T. (2006). Specific language impairment, theory of mind, and visual perspective taking: Evidence for simulation theory and the developmental role of language. *Child Development, 77*(6), 1842-1853.

Fay, W. H., & Schuler, A. L. (1980). *Emerging language in autistic children.* Baltimore: University Park Press.

Fisher, N., Happé, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry*, *46*, 409-419.

Flavell, J. (1977). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.

Fodor, J.A. (1992). The theory of the child's theory of mind. *Cognition, 44,* 283-296.

Frith CD, Frith U. 2000. The physiological basis of theory of mind: functional neuroimaging studies. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ, (Eds). 2000. *Understanding Other Minds*. London/New York: Oxford Univ. Press

Frye, D., & Moore, C. (Eds.) (1991). *Children's theory of mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Frye, D., Zelazo, P.D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10,* 483-528.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia, 38*(1), 11-21.

Gerken, L., Landau, B., & Remez, R. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology, 27,* 204–216.

Gillot, A., Furniss, F., & Walter, A. (2004). Theory of mind ability in children with specific

language impairment. *Child Language Teaching & Therapy*. Arnold Publishers

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16,* 1-15.

Gopnik, A. (1996). Theories and modules; creation myths, developmental realities, and Neurath's boat. In: Carruthers, P., & Smith, P.K (Eds.), Theories of theories of mind (pp. 169-183). Cambridge: Cambridge University Press.

Gopnik, A., & Meltzoff, A. (1993). Imitation, Cultural Learning and the Origins of Theory of Mind. *Behavioral and Brain Sciences, 16*(3), 521-523.

Gopnik, A., Wellman, H.M. (1994). The theory theory. In: L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257-293). New York: Cambridge University Press.

Győri, M., Hahn, N., Várnai, Zs., Sajó, E., Stefanik, K., Balázs, A. (2007). Nem verbális eljárás a hamisvélekedés-tulajdonítás tesztelésére: eredmények tipikusan fejlődő és atipikus fejlődésű gyerekektől. In: Racsmány, M. (Szerk.) *A fejlõdés zavarai és diagnosztikai módszerei.* Akadémiai Kiadó.

Hahn, N., Győri, M., Várnai, Zs., & Sajó, E. (2006). Képességmintázatok a nyelvfejlődési zavarokban: nyelv, társas kogníció, mozgás. In: Rosta Katalin (szerk.): *Egy negyedszázados nevelő-fejlesztő munka tapasztalatai.* Budapest: Logopédiai Kiadó

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Developmental Science, 6*(3), 346-359.

Happé, F. G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J Autism Dev Disord, 24*(2), 129-154.

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Dev, 66*(3), 843-855.

Happé , F., & Loth, E. (2002). 'Theory of mind' and tracking speakers' intentions. *Mind & Language, 17*(1-2), 24-36.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences, 21*, 101-114.

Hocking, J., & Price, C. J. (2009). Dissociating verbal and nonverbal audiovisual object processing. *Brain Lang, 108*(2), 89-96.

Hughes, C., Adlam, A., Happe, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test--retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *J Child Psychol Psychiatry, 41*(4), 483-490.

Jacques, S., & Zelazo, P.D. (2005). Language and development of cognitive flexibility: Implications for theory of mind. In: Astington, J.W. & Baird, J.A. (Eds). *Why language matters for theory of mind.* (pp3-25). New York: Oxford University Press.

Jarrold, C., Boucher, J. & Russell, J. (1997). Language profiles in children with autism: theoretical and methodological implications. *Autism, 1,* 57-76.

Jarrold, C., & Brock, J. (2004). To match or not to match? Methodological issues in autism-related research. *Journal of Autism and Developmental Disorders, 34*(1), 81-86.

Joseph, R. M., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *J Child Psychol Psychiatry, 43*(6), 807-821.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, MA/London: MIT Press.

Karmiloff-Smith, A., Scerif, G., & Ansari, D. (2003). Double dissociations in developmental disorders? Theoretically misconceived, empirically dubious. *Cortex, 39*(1), 161-163.

Kaufman, A.S. (1979). *Intelligent testing with the WISC-R.* New York : Wiley.

Kiss, Sz. (2005). *Elmeolvasás*. Új Mandátum Könyvkiadó.

Kobayashi, C., Glover, G. H., & Temple, E. (2007). Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia, 45*(7), 1522-1532.

Leonard, L. B. (2000). *Children with Specific Language Impairment*. Cambridge: MIT Press.

Leslie, A. M., & Frith, U. (1988). Autistic Childrens Understanding of Seeing, Knowing and Believing. *British Journal of Developmental Psychology, 6*, 315-324.

Lillard, A.S. (1994). Making sense of pretence. In: Lewis, C. & Mitchell, P. (Eds.), *Children's early understanding of mind: Origins and development.* Hillsdale, NJ: Erlbaum.

Lind, S. E., & Bowler, D. M. (2009). Language and Theory of Mind in Autism Spectrum Disorder: The Relationship Between Complement Syntax and False Belief Task Performance. *Journal of Autism and Developmental Disorders, 39*(6), 929-937.

Lohmann, H., Carpenter, M., & Call, J. (2005). Guessing versus choosing - and seeing versus believing - in false belief tasks. *British Journal of Developmental Psychology, 23*, 451-469.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*(4), 1130-1144.

Marcel, A. J. (1993). Slippage in the Unity of Consciousness. *Ciba Foundation Symposia, 174*, 168-186.

Markman, E. M. (1994). Constraints on Word Meaning in Early Language-Acquisition. *Lingua, 92*(1-4), 199-227.

Mayes, L. C., Klin, A., Tercyak, K. P., Jr., Cicchetti, D. V., & Cohen, D. J. (1996). Test-retest reliability for false-belief tasks. *J Child Psychol Psychiatry, 37*(3), 313-319.

Miller, C. A. (2001). False belief understanding in children with specific language impairment. *Journal of Communication Disorders, 34*(1-2), 73-86.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Dev, 78*(2), 622-646.

Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development, 72*(3), 688-690.

Mottron, L. (2004). Matching strategies in cognitive research with individuals with high-functioning autism: Current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders, 34*(1), 19-27.

Muller, U., Zelazo, P.D., Frye, D., & Lieberman, D. (2002). The developmental relations among perspective taking skills, language development, and cognitive flexibility in toddlers. Paper presented at the Jean Piaget Society Conference, Philadelphia.

Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Curr Dir Psychol Sci, 16*(5), 269-274.

Nelson, K. (2005). Language pathways into the community of minds. In: Astington, J.W. & Baird, J.A. (Eds). *Why language matters for theory of mind.* (pp3-25). New York: Oxford University Press.

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory.* Cambridge: Harvard University Press.

Olson, D. (1988). On the origins of beliefs and other intentional states in children. In: Astington, J.W. Harris, P.L., & Olson, D.R. (Eds.), *Developing theories of mind*. New York: Cambridge University Press.

Onishi, K. H., & Bailargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255-258.

Perner, J. (1991). Understanding the representational mind. Cambridge, MA: MIT Press.

Perner, J. (2000). About + belief + counterfactual. In: Mitchell, P., & Riggs, K.J. (Eds.), *Children's reasoning and the mind* (pp. 367-401). Hove, UK: Psychology Press.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the Autistic Childs Theory of Mind - Knowledge, Belief, and Communication. *Child Development, 60*(3), 689-700.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). 2-Year-Olds Difficulty with False Belief - the Case for a Conceptual Deficit. *British Journal of Developmental Psychology, 5*, 125-137.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science, 308*(5719), 214-216.

Perner, J., & Wimmer, H. (1985). John Thinks That Mary Thinks That - Attribution of 2nd-Order Beliefs by 5-Year-Old to 10-Year-Old Children. *Journal of Experimental Child Psychology, 39*(3), 437-471.

Perner, J., Zauner, P., & Sprung, M. (2005). *What does "that" have to do with point of view? Conflicting desires and "want" in German.* In: Astington, J.W. & Baird, J.A. (Eds). *Why language matters for theory of mind.* (pp3-25). New York: Oxford University Press.

Piaget, J. (1967). *The child's conception of the world*. (J. &. A. Tomlinson, Trans.).London : Routledge & Kegan Paul.

Pinker, S. (1999). *Words and Rules*. NY: Basic Books.

Pléh, Cs., Kas, B., Lukács, Á. (2008). A nyelvi fejlődés zavarai. In: Kállai János, Bende István, Karádi Kázmér, Racsmány Mihály (szerk.) *Bevezetés a neuropszichológiába*. Budapest: Medicina. 287-337.

Premack, D., & Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind. *Behavioral and Brain Sciences, 1*(4), 515-526.

Prizant, B. M. (1983). Language-Acquisition and Communicative-Behavior in Autism - toward an Understanding of the Whole of It. *Journal of Speech and Hearing Disorders, 48*(3), 296-307.

Rapin, I & Allen, E. (1987). Developmental dysphasia and autism in preschool children characteristics and subtypes. In: Martin -Martibn-Fletcher- Grunwell-Hall (Eds.), *Proceedings of the first international symposium on SLI and language disorders in children.* London AFASIC, 20-35

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*(1), 12-21.

Riggs, K. J., & Simpson, A. (2005). Young children have difficulty ascribing true beliefs. *Developmental Science, 8*(3), F27-F30.

Rogers, S. J., & Pennington, B. F. (1991). A theoretical approach to the deficits in infantile autism. *Development and Psychopathology, 3,* 137-162.

Roth, D., & Leslie, A.M. (1998). Solving belief problems: Toward a task analysis. *Cognition*, *66,* 1–31.

Ruffman, T., & Keenan, T. R. (1996). The belief-based emotion of surprise: The case for a lag in understanding relative to false belief. *Developmental Psychology, 32*(1), 40-49.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*(3), 734-751.

Scholl, B. J., & Leslie, A. M. (2001). Minds, modules, and meta-analysis. *Child Development, 72*(3), 696-701.

Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., Osanai, H., & Csibra, G. (in press). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology*.

Siegal, M. & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition, 38,* 1–12.

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23*, 117-141.

Slobin, D.I., & Welsh, C. (1973). Elicited imitation as a research tool in developmental psycholinguistics. In: C.A. Ferguson and D.I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart and Winston.

Smith, M., Apperly, I., & White, V. (2003). False belief reasoning and the acquisition of relative clause sentences. *Child Development, 74*(6), 1709-1719.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychol Sci, 18*(7), 587-592.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective.* New York: Oxford University Press, 117–137.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language 17,* 3-23.

Stefanik, K. (2006). *Az iskoláskorra vonatkozó prognósis kérdése autizmusban – a korai képességmintazat és a naiv tudatelméleti vonatkozásu viselkedések szerepe.* Doktori Disszertáció, ELTE, Budapest.

Sullivan, K. & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology, 56,* 135–148.

Sullivan, K., Zaitchik, D., & Tagerflusberg, H. (1994). Preschoolers Can Attribute 2nd-Order Beliefs. *Developmental Psychology, 30*(3), 395-402.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychol Sci, 18*(7), 580-586.

Tager-Flusberg, H. (1992). Autistic Childrens Talk About Psychological States - Deficits in the Early Acquisition of a Theory of Mind. *Child Development, 63*(1), 161-172.

Tager-Flusberg, H. (1997). Language acquisition and theory of mind: contributions from the study of autism. In: L.B. Adamson & M.A. Romski (Eds.), *Research on communication and language disorders: Contributions to theories of language development.* Baltimore, MD: Paul Brookes Publishing.

Tager-Flusberg, H. (2000). Language and Understanding Minds: Connections in Autism In: S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience*. Second Edition. Oxford: Oxford University Press.

Tager-Flusberg, H., & Sullivan, K. (1994). Predicting and Explaining Behavior - a Comparison of Autistic, Mentally-Retarded and Normal-Children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 35*(6), 1059-1075.

Tager-Flusberg, H., & Sullivan, K. (2000). A componential view of theory of mind: evidence from Williams syndrome. *Cognition, 76*(1), 59-89.

Tardif, T., So, C. W. C., & Kaciroti, N. (2007). Language and false belief: Evidence for general, not specific, effects in Cantonese-speaking preschoolers. *Developmental Psychology, 43*(2), 318-340.

Tardif, T., & Wellman, H. M. (2000). Acquisition of mental state language in Mandarin- and Cantonese-speaking children. *Developmental Psychology, 36*(1), 25-43.

Temple, C. (1997). *Developmental Cognitive Neuropsychology.* Oxford: Psychology Press

Tomasello, M. (1995). Joint attention as social cognition. In: Moore, C. & Dunham P.J. (Eds.), *Joint attention: Its origins and role in development (*pp. 103-130). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tomasello, M. (1999). Having intentions, understanding intentions, and understanding communicative intentions. In : Zelazo, P.D., Astington, J.W., & Olson, D.R. (Eds.), Developing theories of intentions (pp. 63-75). Mahwah, NJ: Lawrence Erlbaum Associates.

Tomasello, M. (2000). Culture and cognitive development. *Current Directions in Psychological Science, 9*(2), 37-40.

Tomasello, M. (2003*). Constructing A Language. A Usage-Based Approach.* Cambridge, MA; London, England: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci, 28*(5), 675-691; discussion 691-735.

Tucker, L. (2004). *Specific language impairment and theory-of mind: Is normal language development an essential precursor for on time theory-of-mind development?* Unpublished manuscript, University of Western Australia.

van der Lely, H. (2005). Domain-specific cognitive systems: insight from Grammatical-SLI. *TRENDS in Cognitive Sciences*, 9, 53-59.

Volkmar, F. & Klin A. (2000) Diagnostic Issues in Asperger Syndrome. In Klin, A., Volkmar, F.R., & Sparrow, S.S. (Eds.), *Asperger Syndrome* (p. 25-71). New York: The Guilford Press.

Watson, J. K., Gelman, S. A., & Wellman, H. M. (1998). Young children's understanding of the non-physical nature of thoughts and the physical nature of the brain. *British Journal of Developmental Psychology, 16*, 321-335.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655-684.

Whiten, A. (1991). *Natural theories of mind: Evolution, development and simulation of everyday mindreading.* Oxford: Basil Blackwell.

Wilkinson, K. M. (1998). Profiles of language and communication skills in autism. *Mental Retardation and Developmental Disabilities Research Reviews, 4*(2), 73-79.

Wimmer, H., & Perner, J. (1983). Beliefs About Beliefs - Representation and Constraining Function of Wrong Beliefs in Young Childrens Understanding of Deception. *Cognition, 13*(1), 103-128.

Wing, L. & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children epidemiology and classification. *Journal of Autism and Childhood Schizophrenia, 9,* 11-29.

Zabrucky, K., & Ratner, H. H. (1986). Childrens Comprehension Monitoring and Recall of Inconsistent Stories. *Child Development, 57*(6), 1401-1418.

Ziatas, K., Durkin, K., & Pratt, C. (1998). Belief term development in children with autism, Asperger syndrome, specific language impairment, and normal development: Links to theory of mind development. *Journal of Child Psychology and Psychiatry, 39*(5), 755-763.

Chapter 8. Appendix

Appendix 1. Diagnosis of children with developmental language impairment

	BNO-10 Kód	Diagnózis
1	F80.9	Nem meghatározott zavara a beszéd és a nyelv fejlődésének
2	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
3	F80.1	A kifejező (expresszív) beszéd zavara
4	F80.1	A kifejező (expresszív) beszéd zavara
5	F80.8	Egyéb nyelvi és beszédfejlődési zavar
6	F82.0	A motoros funkció specifikusfejlődésindellenességei
	F80.2	A beszédmegértés (receptív beszéd) zavara
7	F80.1	A kifejező (expresszív) beszéd zavara
	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
8	G24.9 Synd.	Dystonia, k.m.n.
9	F80.2	A beszédmegértés (receptív beszéd) zavara
10	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
11	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
12	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
13	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
14	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
15	F80.2	A beszédmegértés (receptív beszéd) zavara
16	F80.1	A kifejező (expresszív) beszéd zavara
17	F80.8	Egyéb nyelvi és beszédfejlődési zavar
18	F80.9	Nem meghatározott zavara a beszéd és a nyelv fejlődésének
	F80.1	A kifejező (expresszív) beszéd zavara
19	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
20	F80.1	A kifejező (expresszív) beszéd zavara
	F80.2	A beszédmegértés (receptív beszéd) zavara
21	F80.8	Egyéb nyelvi és beszédfejlődési zavar

Appendix 2. Verbal FBT (based on Baron-Cohen et al, 1985)

"Főzés" feladat HAMIS VÉLEKEDÉS
vizsgálatvezető szövege
Látod, ez itt megint Zsuzsi! Csokikrémet főz, és beleteszi ezt a csokit beleteszi a zöld lábosba. Anna közben nézi.
Zsuzsi kicsit főzi a csokit, aztán kimegy valamiért.
Anna ott marad, és kiveszi a csokit a lábosból, és átteszi a másik lábosba. Aztán ő is kimegy.
Zsuzsi nemsokára visszajön, és folytatni szeretné a csokikrém főzést.
Melyik láboshoz megy először?
Miért oda?
Hol van igazából a csoki?
Hová tette Zsuzsi a csokit az elején?
Ez kicsoda? És hogy hívják a másik babát?

"Alvós" feladat IGAZ VÉLEKEDÉS
vizsgálatvezető szövege
Most megint Zsuzsival és Annával fogunk játszani.
Annának van egy macija, és beleteszi ebbe a szekrénybe. Zsuzsi közben nézi.
Anna nagyon álmos, ezért lefekszik aludni. Már el is aludt.
Ezalatt Zsuzsi kiveszi a szekrényből Anna maciját és kicsit játszik vele, majd visszateszi. Ezután elmegy.
Anna nem sokkal később felébred, és játszani szeretne a macijával.
Hol fogja Anna először keresni a maciját?
Miért ott keresi?
Hol van igazából a maci?
Hová tette Anna a macit az elején?
Ez kicsoda? És hogy hívják a másik babát?

"Sally-Anne" feladat HAMIS VÉLEKEDÉS
vizsgálatvezető szövege
Ez Zsuzsi, ez pedig itt Anna.
Zsuzsinak van egy kosara, Annának pedig van egy doboza.
Zsuzsinak van egy golyója is. Beteszi ide, a kosarába, aztán elmegy.
Anna ott marad, és kiveszi Zsuzsi golyóját a kosárból, és átteszi a saját dobozába. Aztán ő is elmegy.
Zsuzsi nemsokára visszajön, és játszani szeretne a golyójával.
Hol fogja Zsuzsi először keresni a golyóját?
Miért ott keresi?
Hol van igazából a golyó?
Hová tette Zsuzsi a golyót az elején?
Ez kicsoda? És hogy hívják a másik babát?

Appendix 3. Complements in wh-question

1. Ez a kislány elment vásárolni reggel. Nagyon sietett haza. Útközben véletlenül elszakította a ruháját egy kerítés kiálló drótjában. Este, amikor ment lefeküdni, azt mondta az anyukájának: „Figyelj, anya, reggel elszakítottam a ruhámat!"

Q1: Mikor mondta az anyukájának, hogy mit szakított el?
Q1B: [csak ha Q1-re rosszul válaszolt]: Mikor mondta el az anyukájának, hogy elszakította a ruháját?
Q2: Mit mondott az anyukájának, mikor szakította el?
Q2B: [csak ha Q2-re rosszul válaszolt]: Mit mondott az anyukájának, mikor szakította el a ruháját?
CQ: Mit csinált a kislány az elején? Mi történt igazából a ruhájával?

2. Ez a néni egyik este vett egy nagy rúd szalámit, a családjának reggelire. Betette a hűtőszekrénybe, és ott is maradt, egész éjjel. Amikor mindenki felébredt, reggel azt mondta a néni: „Tegnap vettem egy nagy rúd szalámit!”

Q1: Mikor mondta a néni, hogy mit vett?
Q1B: [csak ha Q1-re rosszul válaszolt]: Mikor mondta a néni, hogy vett egy nagy rúd szalámit?
Q2: Mit mondott a néni, mikor vett valamit?
Q2B: [csak ha Q2-re rosszul válaszolt]: Mit mondott a néni, mikor vett egy nagy rúd szalámit?
CQ: Tényleg szalámit vett a néni? Igazából mikor vette?

3. Ez a bácsi elindult kirándulni. A szendvicsét reggel a hátizsákjába rakta. Egész nap bandukolt az erdben. Egyszer csak megéhezett, és enni akart. Este volt már, de azt gondolta, hogy reggel a szendvicsét a hátizsákjába rakta.

Q1: Mikor gondolt arra, hogy hová rakta a szendvicsét?
Q1B: [csak ha Q1-re rosszul válaszolt]: Mikor gondolta, hogy a szendvicsét a táskájába rakta?
Q2: Mit gondolt a bácsi, mikor rakta el a szendvicsét?
Q2B: [csak ha Q2-re rosszul válaszolt]: Mit gondolt a bácsi, mikor rakta el a szendvicsét a táskájába?
CQ: Hová indult a bácsi? Igazából hová rakta a szendvicsét?

4. Ez a macska szeretné megfogni ezt az egeret. Látta, hogy az egér este bement az egérlyukba. Másnap visszajött a macska, hogy megvárja, míg el%jön az egér. Álldogált ott reggel, és azt gondolta, este az egérlyukba ment be az egér.

Q1: Mikor gondolt arra a macska, hogy hová ment be az egér?
Q1B: [csak ha Q1-re rosszul válaszolt]: Mikor gondolt arra a macska, hogy az egérlyukba ment be az egér?
Q2: Mit gondolt a macska, mikor ment be az egér?

Q2B: [csak ha Q2-re rosszul válaszolt]: Mit gondolt a macska, mikor ment be az egér?
CQ: Igazából hová ment be az egér? Mit akart a macska?

5. Ez a fiú délben, az iskolában almát evett ebédre. Később hazament, és otthon játszott. Este aztán az anyukájának azt mondta a fiú: „Képzeld, ma délben csokoládét ettem!"

Q1: Mikor mondta a fiú az anyukájának, hogy mit evett?
Q1B: [csak ha Q1-re rosszul válaszolt]: Mikor mondta az anyukájának, hogy csokoládét evett?
Q2: Mit mondott az anyukájának, mikor evett?
Q2B: [csak ha Q2-re rosszul válaszolt]: Mit mondott az anyukájának, hogy mikor evett csokoládét?
CQ: Igazából mit ebédelt a fiú az iskolában?

6. Ez a kisfiú hazafelé menet az iskolából délután a buszon elvesztette a táskáját. Nagyon szomorú lett emiatt. Amikor az anyukája hazajött este, a kisfiú azt mondta neki: „Délután a villamoson elvesztettem a táskámat."

Q1: Mikor mondta a kisfiú az anyukájának, hogy hol vesztette el a táskáját?
Q1B: [csak ha Q1-re rosszul válaszolt]: Mikor mondta a kisfiú az anyukájának, hogy a villamoson vesztette el a táskáját?
Q2: Mit mondott a kisfiú, mikor vesztette el a táskáját?
Q2B: [csak ha Q2-re rosszul válaszolt]: Mit mondott a kisfiú, mikor vesztette el a táskáját a villamoson?
CQ: Igazából hol vesztette el a táskáját? Honnét jött akkor?

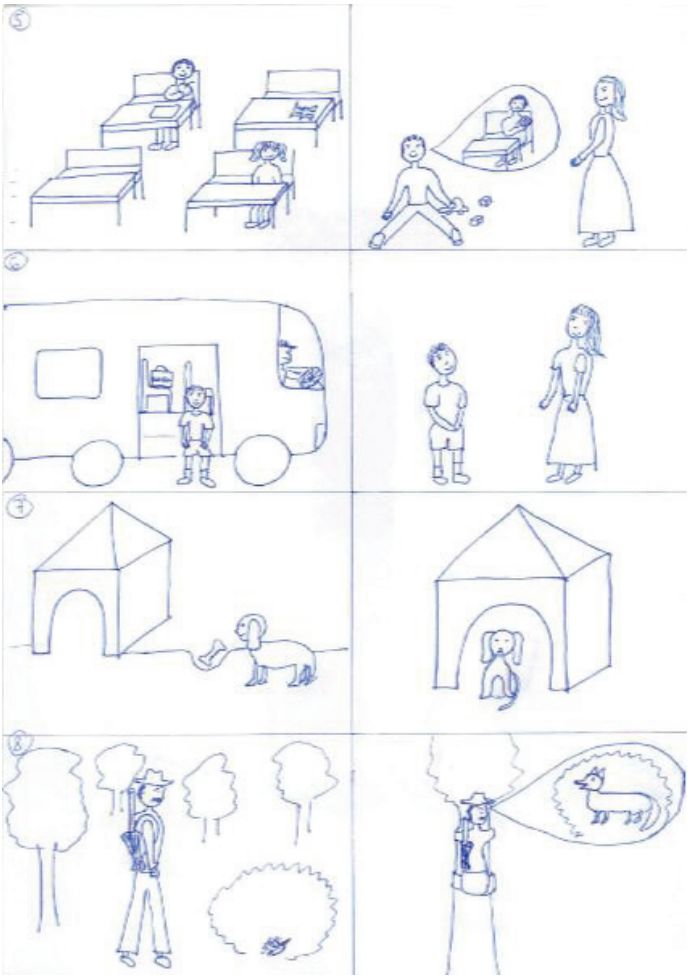
7. Ez a kutyus talált egy csontot az udvaron. Este elásta magának a kutyaház mellett. Éjjel jó nagyot aludt, és aztán nagyon éhesen ébredt. Reggel azt gondolta magában, este elásott egy csontot a kútnál!

Q1: Mikor gondolta magában a kutyus, hogy hol ásta el a csontot?
Q1B: [csak ha Q1-re rosszul válaszolt]: Mikor gondolta magában a kutyus, hogy a kútnál

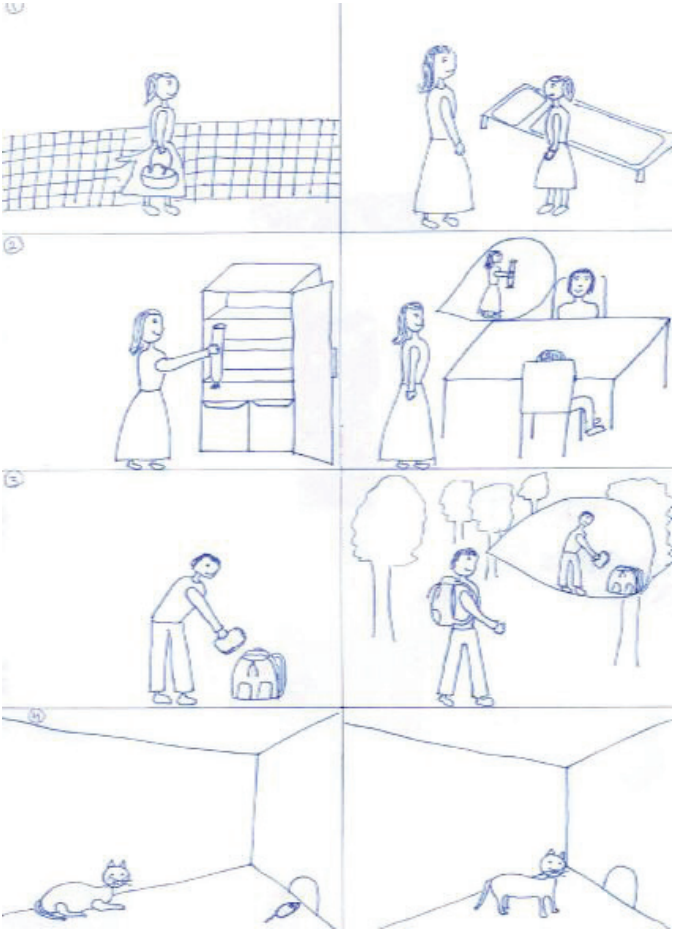
ásta el a csontot?
Q2: Mit gondolt a kutyus, mikor ásta el a csontot?
Q2B: [csak ha Q2-re rosszul válaszolt]: Mit gondolt a kutyus, mikor ásta el a csontot a kútnál?
CQ: Hogyan szerezte a kutyus a csontot? Hol ásta el?

8. Ez a vadász sétált az erdben. Ahogy elsétált egy bokor mellett, a bokorban megmozdult egy süni. A vadász nem látta, nagyon megijedt, és ijedtében felmászott egy fára. A fa tetején a vadász azt gondolta, hogy egy farkas mozog a bokorban.

Q1: Hol gondolta a vadász, hogy mi mozog?
Q1B: [csak ha Q1-re rosszul válaszolt]: Hol gondolta a vadász, hogy farkas mozog?
Q2: Mit gondolt a vadász, hol mozog valami?
Q2B: [csak ha Q2-re rosszul válaszolt]: Mit gondolt a vadász, hol mozog a farkas?
CQ: Igazából mi mozgott a bokorban? Mit csinált a vadász az erdőbe?



74



Appendix 4. Word-learning complement task

Instrukció: Most történeteket fogok neked elmesélni, és hozzá képeket fogok mutatni. Kérlek, hogy jól nézd meg a képeket, figyelmesen nézd meg minden részletüket, mert utána kérdéseket fogok feltenni a képekről.

1. Ő azt mondta a kutyájának, hogy egy KAMPA van a fűben, és tényleg egy KAMPA volt a fűben.

	kérdés	válasz	hányadik	OK?
1	Mit mondott a kutyájának?			
Tesztelési helyzet: a 2 tárgy képének megmutatása				
2	Mutasd meg a KAMPÁt!			
3	Melyik a KAMPA: ez itt a KAMPA? (random módon felemelve KAMPÁT vagy másik tárgy képét)			
4	Ez a KAMPA? (felemelve az előbb nem mutatott tárgy képét)			
5	Add ide a KAMPÁt!			
6	**Melyik volt igazából a fűben? (memória)**			
7	**Melyik volt az úton? (memória)**			
8	Melyik tetszett jobban?			

2. Ő azt mondta a lánynak, hogy egy könyv van a dobozban, de igazából egy LUTI volt benne.

	kérdés	válasz	hányadik	OK?
1	Mit mondott a lánynak?			
Tesztelési helyzet: a 2 tárgy képének megmutatása				
2	Mutasd meg a LUTIt!			
3	Melyik a LUTI: ez itt a LUTI? (random módon felemelve LUTIT vagy másik tárgy képét)			
4	Ez a LUTI? (felemelve az előbb nem mutatott tárgy képét)			
5	Add ide a LUTIt?			
6	**Melyik volt igazából a dobozban? (memória)**			
7	**Melyik volt a földön? (memória)**			
8	Melyik tetszett jobban?			

3. Ő azt mondta a kislánynak, hogy egy játék van a kezében, de igazából egy TIMA volt benne.

	kérdés	válasz	hányadik	OK?
1	Mit mondott a kislánynak?			
Tesztelési helyzet: a 2 tárgy képének megmutatása				
2	Mutasd meg a TIMÁt!			
3	Melyik a TIMA: ez itt a TIMA? (random módon felemelve TIMÁT vagy másik tárgy képét)			
4	Ez a TIMA? (felemelve az előbb nem mutatott tárgy képét)			
5	Add ide a TIMÁt?			
6	**Melyik volt igazából a kezében? (memória)**			
7	**Melyik volt az úton? (memória)**			
8	Melyik tetszett jobban?			

4. Ő azt mondta a testvérének, hogy egy pulóver van a táskájában, de igazából egy PERI volt benne.

	kérdés	válasz	hányadik	OK?
1	Mit mondott a testvérének?			
Tesztelési helyzet: a 2 tárgy képének megmutatása				
2	Mutasd meg a PERIt!			
3	Melyik a PERI: ez itt a PERI? (random módon felemelve PERIT vagy másik tárgy képét)			
4	Ez a PERI? (felemelve az előbb nem mutatott tárgy képét)			
5	Add ide a PERIt?			
6	**Melyik volt igazából a táskában? (memória)**			
7	**Melyik volt az úton? (memória)**			
8	Melyik tetszett jobban?			